ELSEVIER

Journal of Hydrology 192 (1997) 81–103

Journal of Hydrology

Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling

Thian Yew Gan*, Enoch M. Dlamini[1], Getu Fana Biftu

Department of Civil Engineering, University of Alberta, Edmonton, Alta. T6G 2G7, Canada

Received 3 November 1995; revised 29 March 1996; accepted 2 April 1996

Abstract

Three medium sized, dry catchments located in Africa and USA were modeled with four or five conceptual rainfall–runoff (CRR) models of different complexity. The models were the Pitman model of South Africa (16 parameters), the Sacramento model of USA (21 parameters), the NAM model of Europe (15 parameters), the Xinanjiang model of China (15 parameters), and the SMAR model of Ireland (nine parameters). Between these models, the Xinanjiang model has been consistently doing better mainly because it is the only model that considers the non-uniform distribution of runoff producing areas to simulate the runoff. On the whole, it seems that standard, good quality hydrologic data can still support modeling of dry catchments with traditional CRR models. The model performance depends more on the model structure, the objective function used in automatic calibration, and data quality, than on model complexity or calibration data length. For relatively dry catchments such as the great Usuthu catchment, wet years should be preferred over dry years for the calibration data. © 1997 Elsevier Science B.V.

1. Introduction

Catchments with streamflow/rainfall ratios of about 0.2 or less (henceforth referred to as dry catchments) are generally more difficult to model than wet catchments or catchments with relatively high streamflow/rainfall ratios because the hydrologic processes of the former are more complex and variable than the latter. The dry spells of dry catchments are usually marked with minimal streamflow supported mostly by groundwater but sometimes the transition to wet seasons can happen quite drastically. The arid climate of many dry catchments often results in patchy and sparse vegetation, particularly on high spots.

* Corresponding author.
[1]Present address: Water Resources Division, Box 58, Mbabane, Swaziland.

PII S0022-1694(96)03114-9

One would expect the hydrologic processes from barren high spots to be quite different from valleys or areas covered with vegetation, as well as the runoff contributions from these two areas. Even though the variable source concept of Hewlett (1982), which means that only parts of a catchment contribute flow to the stream, also applies to wet catchments, the more abrupt and heterogeneous changes taking place in dry catchments cause the hydrologic data of these catchments to be generally noisy or highly variable, unrepresentative, or even erroneous.

The majority of the traditional conceptual rainfall–runoff (CRR) models have been built for temperate or wet catchments where modeling is more straightforward, or where the hydrologic responses only involve a subset of the processes which occur in dry climatic regions. Despite that, with the limited amount of information contained in the standard, hydrologic data, Jakeman and Hornberger (1993) found that the permissible model complexity for temperate catchments is only about half a dozen parameters. If that were the case, many traditional CRR models are perhaps more complex than what the standard data can support. Because the physics governing the catchment-scale rainfall–runoff transformation process is highly complex, CRR models that incorporate major hydrologic processes often have more than half a dozen of parameters. Table 1 shows that the CRR models considered here have between nine and 21 parameters.

The question of over-parameterization is a major problem to be dealt with in rainfall–runoff modeling. Beven (1989) commented that three to five parameters should be sufficient to reproduce most of the information in a standard data set. Hornberger et al. (1985) and Loague and Freeze (1985) found more complex models did worse than simple models. With ever increasing computing power, and the availability of more and more powerful optimization algorithms, the danger of over-parameterization may be growing ever larger. What should be the generally allowable level of model complexity for dry catchments, given that such catchments undergo more complex and a wider range of hydrologic processes than temperate or wet catchments, and have only similar or even worse data sets for model calibrations? Will the majority of traditional CRR models built for temperate catchments suffer from over-parameterization if applied to dry catchments? What impact will the choice of calibration data have on calibrating traditional CRR models to dry catchments, given that the standard rainfall–runoff data only contain the input–output information, not the information on the rainfall–runoff transformation process? Can a powerful, global optimization algorithm such as SCE-UA (shuffle complex evolution method, see Section 5.1 increase the chances of successfully applying more complex models to dry catchments?

2. Research objectives

With the above questions as background, this study has three research objectives on the hydrologic modeling of dry catchments:

1. What is the allowable level of CRR model complexity and what are the effects of model structures?
2. What is the influence of the choice of objective functions on the automatic calibration of CRR models?

3. What are the effects of the choice of wet, average, and dry year calibration data and data length on model calibration?

To fulfil the above objectives, this study compared the performance of five CRR models of different structures and complexity on three catchments, two located in Africa and one in the United States. Two African catchments and an African CRR model (Pitman) were chosen for this study partly because there have been relatively few hydrologic modeling studies conducted in Africa. The comparisons between an African model and CRR models built in other continents should be of interest to many hydrologic communities, particularly the Africans. Therefore the discussion of results is centered on the African catchments.

3. Hydrologic regimes of dry catchments

In this section, the hydrologic regimes of three medium sized, dry catchments are briefly described. Although all three catchments have an average annual precipitation slightly less than 1000 mm, they are classified as dry catchments because of the following characteristics found in these catchments: relatively low streamflow/rainfall ratios of about 0.2 or less, generally long and severe dry seasons, and a drastic change from dry to wet seasons (see Fig. 1).

3.1. Great Usuthu catchment of Swaziland

The great Usuthu river basin of Swaziland has a basin area of 2682 km^2. The streamflow has 4 months of high flow from January to April, and from April onwards declines continuously. The vegetation of the catchment is predominantly open grassveld, most of which is covered with pine plantations. The catchment has experienced a changing climate in the past several decades: the wet 50s, dry 60s, and wet 70s (Fig. 2). The catchment has a mean annual precipitation of 900 mm, a mean annual potential evapotranspiration of 1270 mm, and a mean annual runoff of 154 mm (streamflow/rainfall ratio = 0.16).

3.2. Ihimbu catchment of Tanzania

Tanzania has a tropical climate with a long dry season of 6 to 9 months. The Ihimbu catchment of Tanzania, with an area of 2480 km^2 covered mainly with savannah woodland and deteriorated forest, is drained by the Little Ruaha River. The catchment has a mean daily rainfall of 2.66 mm day^{-1}, a mean daily potential evaporation of 4.06 mm day^{-1}, and a mean daily discharge of 0.61 mm day^{-1} (streamflow/rainfall ratio = 0.23).

3.3. Bird Creek of United States

The Bird Creek of United States has a catchment area of 2344 km^2. The catchment also has a low streamflow/rainfall ratio (= 0.23) but it differs from the African counterparts

Table 1
Characteristics of the model structures of the Pitman (PTM), Sacramento (SMA), NAM, Xinanjiang (XNJ), and SMAR models

Classification	PTM	SMA	NAM	SMAR	XNJ
Origin	South Africa (Africa)	USA (North America)	Denmark (Europe)	Ireland (Europe)	China (Asia)
Input	Rainfall Potential or pan evaporation	Precipitation Potential evaporation	Rainfall Potential evaporation Snow (optional)	Rainfall Potential evaporation	Rainfall Potential evaporation
Modeling time steps	Daily and monthly	1 h, 6 h, and daily	Daily	Daily	Daily
No. of model parameters	16	21	15	9	15
No. of optimized parameters in this study	9	13	13	6	15
No. of soil moisture zones	1	2	2	5 maximum	3
No. of conceptual storages	3	5	4	3	4
Actual evaporation	Linear relationship with soil water storage	PET[a] multiplied by linear ratio of state versus capacity of tension storages and upper zone free water storage	Potential rate multiplied by the relative water content in the lower zone	Potential rate from upper storage and a coefficient multiplied by the potential rate from lower storage	Potential rate from upper storage and a coefficient multiplied by the potential rate from lower storage
Types of conceptual storages	(i) Interception storage (ii) Soil moisture storage (iii) Groundwater storage	(i) Upper zone tension water storage (ii) Upper zone free water storage (iii) Lower zone tension water storage (iv) Lower zone primary water storage (v) Lower zone supplementary water storage	(i) Snow storage (optional) (ii) Surface water storage (iii) Lower zone soil moisture storage (iv) Upper groundwater storage (v) Lower zone groundwater storage	(i) Surface storage (ii) Maximum five layers of soil moisture storage (iii) Groundwater storage	(i) Upper zone tension water storage (ii) Lower zone tension water storage (iii) Deep zone tension water storage (iv) Free water storage

Types of flows	(i) Direct overland flow (ii) Horton overland flow (iii) Saturation overland flow (iv) Groundwater flow (baseflow)	(i) Direct overland flow (ii) Saturation overland flow (iii) Interflow (iv) Groundwater flow (baseflow)	(i) Direct overland flow (ii) Saturation overland flow (iii) Interflow (iv) Groundwater flow (baseflow)	(i) Direct overland flow (ii) Saturation overland flow (iii) Interflow (iv) Groundwater flow (baseflow)	(i) Direct overland flow (ii) Saturation overland flow (iii) Interflow (iv) Groundwater flow (baseflow)
Sub-basin routing	Separate simple lag functions for groundwater and surface flow	Unit hydrograph for all flows	Separate linear reservoirs for overland flow and interflow upper baseflow and lower baseflow	Cascade of reservoirs (Nash method)	Unit hydrograph for surface flow and linear reservoir for baseflow
Basin routing	Muskingum equation for surface flow	Muskingum equation and variable lag for lag basins only	Linear reservoir	Linear reservoir	Muskingum equation or Nash method

[a] PET, potential evaporation.

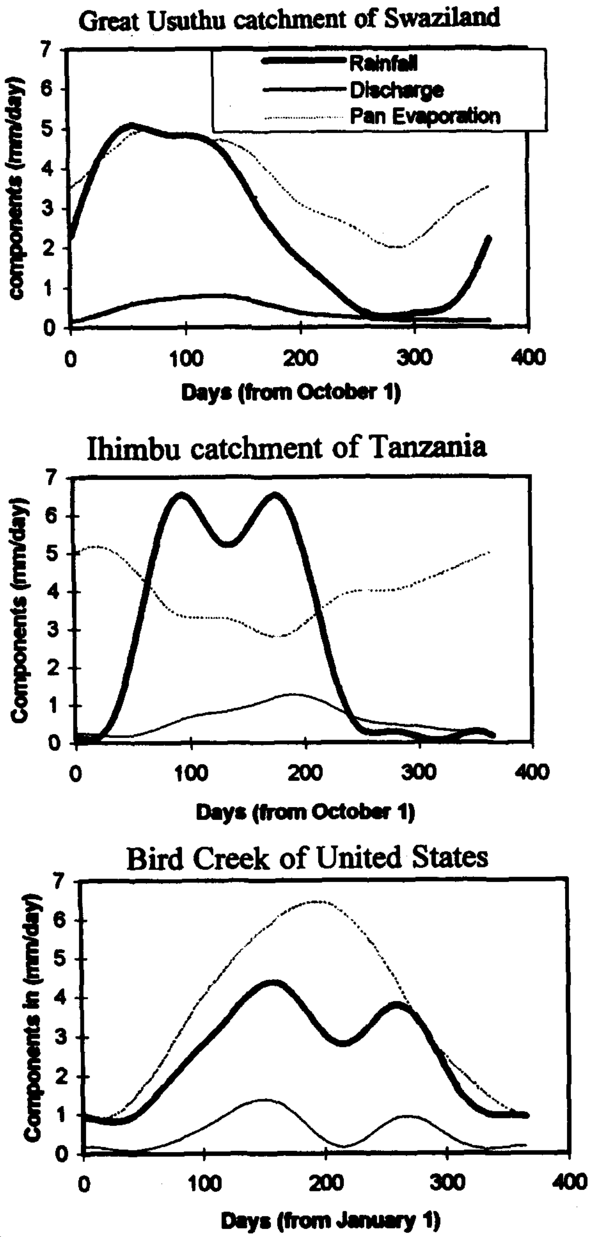

Fig. 1. Plots showing daily averages of every calendar day for the rainfall, discharge, and the pan evaporation in mm day^{-1} for the two African and one USA test catchments.

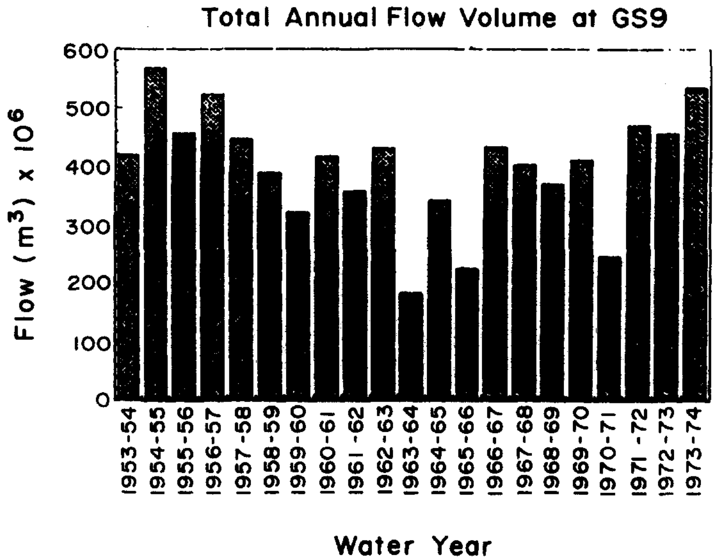

Fig. 2. Total annual flow volume at the GS9 station of the Great Usuthu catchment.

mainly in terms of lower potential evaporation during winter months (due to lower winter temperature) and a smoother transition between wet and dry seasons. It has a mean daily rainfall of 2.66 mm day^{-1}, a mean daily potential evaporation of 3.58 mm day^{-1}, and a mean daily discharge of 0.61 mm day^{-1}.

4. Conceptual rainfall–runoff models

For the past several decades, using the primary approach of transforming rainfall (model input) to streamflow (model output) through a number of interconnected mathematical functions (each representing a certain component of the hydrologic cycle), a wide range of CRR models has been developed. To select a model for a specific application, often three major issues are involved. Firstly, the conceptual base of the model should 'capture' the major hydrological processes of the catchment. A calibrated model may reasonably reproduce observed patterns of streamflow at the calibration stage even though the basin's major hydrologic processes can be different from those assumed in the model. The conceptual base of a CRR model is also linked to its level of complexity which should depend on the amount of detail desired and the types of data available. With an increase in computing power, there is a tendency to apply complex CRR models even though an unwarranted increase in model complexity could introduce the problem of over-parameterization (Beven, 1989, p. 159). By comparing the performance of five CRR models of different complexity on three dry catchments, we can gain some insight into the level of model complexity suitable for dry catchments.

The second issue is whether the time steps used in the model are sufficiently small to represent the rates of change of the process. This problem can be particularly important in arid regions where the rainfall intensity can change quickly over relatively short time scales (Hughes and Herald, 1987), resulting in a sharp rise in the surface runoff. For the three test catchments of slightly over 2000 km^2, daily time steps are generally adequate even though peak flows may require sampling intervals shorter than daily ones. Unfortunately, hydrologic data available in Africa seldom have time steps less than a daily one. The third issue, which concerns model calibration, is given in Section Section 5.

The five conceptual rainfall–runoff models selected for this study were: (i) the Pitman model (PTM) of South Africa, (ii) the Sacramento model (SMA) of USA, (iii) the NAM model of Denmark, (iv) the Xinanjiang model (XNJ) of China, and (v) the SMAR model of Ireland. Between these models, SMAR is the simplest (nine parameters), PTM (16 parameters), XNJ (15 parameters), and NAM (15 parameters) are in between, and SMA is the most complicated (21 parameters). None of these models was designed specifically for dry catchments. Table 1 summarizes various components of these five models and the actual number of parameters calibrated for each model. To ensure proper calibration, one of the latest global optimization algorithms, the shuffle complex evolution algorithm (see Section 5.1, was used in this study.

The Pitman model (Pitman, 1976) has been used successfully throughout Southern Africa, in countries such as Botswana, Lesotho, and Swaziland (Pitman et al., 1981). The model has three conceptual storages: interception, soil moisture, and the groundwater storages. The model produces direct and saturation overland flows when the soil moisture storage is full. These two flow components are routed separately from the groundwater component which is produced by the groundwater storage. Model inputs are rainfall and potential evaporation. The model takes in both monthly and daily rainfall. If the daily rainfall data are incomplete, the monthly data are used to adjust the daily rainfall.

The Sacramento model (SMA), which is the US National Weather Service model for operational river forecasting (Burnash et al., 1973), is one of the most studied CRR models in recent years (e.g. Hendrickson et al., 1988; Gan and Burges, 1990; Sorooshian et al., 1993). The "Nedbor-Afstromnings model," NAM, (Danish Hydraulic Institute, 1982), operates by continuously accounting for moisture content in five mutually interrelated storages: snow, surface, lower zone, upper and lower groundwater storages. The Xinanjiang model (XNJ), developed in 1973, has been used to forecast flows to the Xinanjiang reservoir (Zhao et al., 1980). It was built to forecast floods in large basins in humid regions where the infiltration rate is high, making surface runoff small and interflow or subsurface flow high (leading to a relatively flat hydrograph). The soil moisture accounting and routing, SMAR, model (O'Connell et al., 1970; Kachroo, 1992) has only nine parameters, and it is the simplest model tested in this study.

5. Calibration of CRR models on dry catchments

Model calibration is a major aspect of hydrologic modeling. Conceptually realistic models can produce erroneous results if they are not properly calibrated. Model calibration is either done manually or by a combination of manual effort and automatic procedures (Gan, 1988). The manual calibration stage can be minor or major. It can either involve assigning initial values to parameters which are then optimized entirely by an automatic procedure or major calibrations are done manually while an optimization algorithm is only used for fine-tuning some selected parameters. Whether a model is calibrated by an automatic algorithm or manually, past studies indicate that it is very difficult, if not impossible, to obtain a unique set of optimal parameters for a CRR model (e.g. Gupta and Sorooshian, 1985; Gan and Biftu, 1996). Factors contributing to the dilemma of model calibration include calibration data with limited information, data with measurement errors, spatial variability of rainfall or catchment properties poorly represented by point measurements, and CRR models suffer from model structure and parameter identifiability problems.

When an automatic calibration procedure is used, the final model parameters derived depend primarily on four elements: (i) optimization algorithm, (ii) objective function, (iii) calibration data, and (iv) model structure and identifiability of model parameters. Each element exerts a certain degree of influence on the final outcome of model calibration. The focus of this paper is on the second and third elements.

5.1. *Shuffle complex evolution method*

The shuffle complex evolution method (SCE-UA) of Duan et al. (1992) is a global optimization method classified as probabilistic because it evaluates the objective function at randomly spaced points in the feasible parameter space. SCE-UA was used in this study because it represents a synthesis of the best features of several methods. It combines the strength of the Simplex procedure of Nelder and Mead (1965) with the concept of a controlled random search, a systematic evolution of points in the direction of global improvement, competitive evolution (Holland, 1975), and complex shuffling. Duan et al. (1992) compared the independent global search in the feasible space without sharing information by giving a number of competent people a difficult problem to solve without conferring with each other. They believed that it would be better for the people to first work independently (individually or in small groups), and get together now and then to share information about their progress. They incorporated this feature in SCE-UA.

5.2. *Objective functions*

The objective function used in the generation of the response surface (objective criteria) is crucial in the automatic calibration process. In this study, the maximum likelihood for the heteroscedastic error, HMLE (which assumes the presence of an uncorrelated inhomogeneous variance error; Sorooshian and Gupta, 1983), and a simple daily root mean square objective function, DRMS ($= 1/n\sqrt{\mathrm{SLS}}$, where n is the number of data points, SLS = simple least square), were tested. DRMS, which assumes the presence of a Gaussian, independent homogeneous variance error, does not involve data transformation while HMLE transforms data to the logarithmic space. Three stopping criteria were used to terminate SCE-UA. The search was either stopped after 20 shuffling iterations, or if the change in objective function and the change in parameter values were both less than 0.0001, or if the number of iterations was greater than 50 000.

5.3. Strategy to test the effects of data variability and data length

Ideally, there should be 3 to 5 years of data that include average, wet, and dry years so that the data encompass a sufficient range of hydrologic events to activate all the model parameters during calibration (as shown in Section 6.4). Using 21 years of data from the Great Usuthu catchment, the effects of data variability are investigated. All five models were calibrated with wet years and verified with dry years and vice versa, or calibrated and verified with mixed year data, for example, mixture of wet, dry, and average years. The effects of data lengths were studied by calibrating each model with 2, 5, and 10 years of data. Four sets of tests were done.

In Test I all conceptual rainfall–runoff models were calibrated with 5 relatively wet years (1953 to 1958) and the calibrated models verified with 5 relatively dry years (1963 to 1968). The SCE-UA optimization algorithm based on the DRMS objective function was used to calibrate all five models. Then the objective function was changed to HMLE, keeping everything else the same. Test II is the same as Test I except that 2 years of data were used instead of 4. Test III, which was used to investigate the effect of dry versus wet calibration data, was a repetition of Test II except that the calibration was carried out using 2 relatively dry years of data (1963 to 1965). In Tests I and II, models calibrated from 5 and 2 years of data were also verified with 11 years of mixed (wet, dry, and average) data (1963–1974). Finally, in Test IV, models calibrated with 10 years of mixed data (1953–1963) were verified with 11 years of data (1963–1974). The strategies for these tests are summarized in Table 2 and Table 3 which are results for the Great Usuthu catchment using DRMS and HMLE, respectively. Results for the other catchments are given in Table 4 and Table 5.

6. Discussions of results

Discussions in Section 6.1 to Section 6.4 are mostly based on the Great Usuthu catchment. For the Ihimbu catchment and Bird Creek, the results are discussed in Section 6.5 and Section 6.6, respectively. Three statistical indices selected to compare the performances of the models calibrated with the three catchments are the root mean square error (RMSE), bias (BIAS), and the coefficient of efficiency or Nash–Sutcliffe coefficient (E_f) (Nash and Sutcliffe, 1970). An E_f of 100% means a perfect prediction while a negative E_f means that the model's simulated value is worse than simply using the observed mean. The subsequent discussions are mainly given in terms of E_f, instead of all three statistics. A careful inspection of Table 2 and Table 3 reveals that the three statistics generally gave similar results. Besides statistics, selected plots of runoff hydrographs are included in Fig. 3 and Fig. 4.

6.1. Model comparisons based on the Great Usuthu catchment

Table 2 compares the calibration and verification runs of PTM, SMA, NAM, XNJ, and SMAR models using SCE-UA as the optimization method, DRMS as the objective function, and the same data sets. Table 3 is similar to Table 2 except the objective function used is HMLE and the models compared are PTM and SMA only. On the whole, Table 2 and Table 3 clearly indicate that the DRMS objective function produces better statistical indices (which does not necessarily mean better results) than HMLE. Only in one instance (Test I for PTM) did the reverse happen. Example plots of Fig. 3 and Fig. 4 reveal that DRMS produces better statistical measures because under DRMS high flows are more accurately simulated whereas under HMLE low flows are more accurately simulated. When the model performance is assessed in terms of statistics, DRMS naturally appears to be superior over HMLE since high flows carry more weight than low flows. Given that the results between the tables are similar to each other, further discussions are mainly based on Table 2.

In many calibration runs, XNJ generally performed slightly better than the other four models, NAM and PTM are comparable to each other, and SMAR and SMA performed less satisfactorily (Table 2). At the calibration stage, a good match between simulated and observed flows does not necessarily mean that a proper calibration has been achieved. To truly assess a calibration, verification runs were based on data sets independent of those used during calibration. For example, for the calibration run in Test IV, the E_f values for PTM, XNJ, NAM, SMAR, and SMA are 73.5%, 74.8%, 74.3%, 63.4%, and 69.1%, respectively. For the verification runs in Test IV, the corresponding E_f values are 70.8%, 68.7%, 62.5%, 61.1%, and 60.5%, respectively. These E_f values show that the calibrations achieved are mainly moderate. For the calibration and verification runs in Test II, the drops in E_f are comparable to that in Test IV. Table 2 shows that moving from the calibration to the verification runs, the drop in model performance (in terms of E_f) ranges from less than 1% to over 25%. Given that the E_f obtained can vary over a wide range for the Usuthu catchment (and also for the other two dry catchments), the success of a model calibration likely depends heavily on the calibration data and the model used. It seems that dry catchments are harder to calibrate than wet or temperate catchments (Gan and Biftu, 1996) and the outcome of calibrating dry catchments can be unpredictable.

6.1.1. Sacramento model (SMA)

Even though SMA is a relatively complex model, its performance was poor, and at times even poorer than SMAR, the simplest of the five models. The poor performance of SMA can be partly attributed to its non-standard unit hydrograph called unitgraph (Brazil, 1980), which was used to convert the computed channel inflow into the catchment outflow. This unitgraph more or less matches the traditional unit hydrograph if surface runoff is the dominant runoff. The difference between the SMA unitgraph and the traditional unit hydrograph increases as the flow regime becomes more and more dominated by mixed flow instead of surface flow. Since the catchments tested have a short wet season and a long dry season, it is unlikely that one set of unit hydrograph ordinates will work well.

In this study, the SMA unitgraph ordinates were first derived through model calibration. Because of parameter interactions and the presence of two distinct seasons, it was not possible to get a realistic set of unitgraph ordinates from this approach. A separate attempt was made by regressing streamflow data (dependent variable) with rainfall data (independent variable). The performance of SMA based on the unitgraph ordinates derived from this approach turned out to be worse than before. Perhaps as a more plausible approach, two sets of unitgraphs, one for routing low flows and one for routing high flows, will improve SMA's performance. Because complex models may have one or more components that do not function properly under certain climatic conditions, such as SMA's unitgraph, complex models may produce poorer results than simpler models. Similar observations were also found by Loague and Freeze (1985) and Hornberger et al. (1985).

6.1.2. Pitman model (PTM)

PTM was built for African catchments but not specifically for dry catchments. It has three features not found in the other four models. Firstly, only PTM is specifically designed to simulate Horton overland flow while other models only simulate direct and saturation overland flows. Even with this feature, PTM also has problems simulating high flows. To simulate high flows accurately it is important to have accurate data of time steps finer than daily ones since major storms often undergo significant changes within hours. Since the data available are only daily averages, and PTM was operated at daily time steps, high flows could not be simulated accurately.

Secondly, PTM breaks down daily rainfall depths to hourly increments according to a regression of the form, duration (h) = $\alpha + \beta$(rainfall) (mm), where α and β are linear regression parameters. Pitman derived the α and β parameters for Pretoria, South Africa. This approach has two potential pitfalls. The relationship between storm duration and rainfall depths for a geographical location may not necessarily be linear and hourly rainfall data are needed to derive α and β for that location. Since no actual hourly rainfall data were available for the Great Usuthu or Ihimbu catchments, the α and β values derived for Pretoria were also used for the Great Usuthu catchment. For Ihimbu, it was necessary to derive new α and β values through model calibration.

Thirdly, PTM uses a variable recession constant to compute the baseflow while other models use fixed recession constants. In view of the prolonged dry spell of dry catchments, the former is probably more realistic than the latter. Even with this additional feature, PTM could not do better than XNJ, or NAM in the low flows (Fig. 3 and Fig. 4). Perhaps the approach that PTM used to convert the daily rainfall to hourly increments was not applied properly since there was no hourly rainfall data to accurately determine the values of α and β.

6.1.3. Xinanjiang model (XNJ)

XNJ has been doing marginally better than other models even though XNJ was built for large catchments in humid regions with rich vegetation, well-developed soil zone, low surface runoff, and high interflow. This is likely because while other models assume a uniform distribution in soil moisture in the catchment, XNJ considers a non-uniform spatial distribution of soil moisture deficit and storage capacity in the catchment. By describing these two variables according to a non-uniform spatial distribution, the runoff producing area is also simulated in terms of a non-uniform distribution. Since the runoff producing areas of dry catchments are probably more unevenly distributed than humid catchments, and as the only model that takes this into consideration, XNJ consistently did better than other models in all three catchments. However, in simulating the evaporation, XNJ assumes the soil moisture deficit and the storage capacity to be uniformly distributed, which is inconsistent with the assumption used to simulate runoff.

Table 2

Summary statistics of the calibration and verification runs for Pitman, Sacramento, NAM, Xinanjiang, and SMAR models applied to the Great Usuthu catchment of Swaziland. All the models were calibrated using SCE-UA and the DRMS objective function

Run no.	Data type	Test period (C = calibration, V = verification)	Mean flow (mm day^{-1})	Pitman model (PTM)			Sacramento model (SMA)			NAM model (NAM)			Xinanjiang model (XNJ)			SMAR model (SMAR)		
				E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)
I	5-year wet	C (1953–1958)	0.505	65.8	0.285	–20.15	68.5	0.272	–20.95	76.4	0.232	2.9	78.5	0.221	–5.3	65.6	0.281	–0.9
	5-year dry	V (1963–1968)	0.326	64.6	0.268	–3.83	74.6	0.228	1.39	73.2	0.249	24.3	76.6	0.232	20.4	65.6	0.279	–0.4
	11-year mixed	V (1963–1974)	0.381	58.4	0.319	–3.87	59.6	0.315	–3.06	61.0	0.318	26.5	64.2	0.305	16.1	58.3	0.329	20.4
II	2-year wet	C (1954–1956)	0.542	89.1	0.197	–0.55	85.4	0.230	–12.07	88.5	0.202	–0.9	91.5	0.176	–5.2	87.8	0.212	5.4
	2-year dry	V (1963–1965)	0.270	81.1	0.152	20.40	77.1	0.167	9.32	84.6	0.170	18.2	83.3	0.176	–15.1	86.1	0.161	–33.0
	11-year mixed	V (1963–1974)	0.381	66.2	0.288	19.75	60.2	0.313	–3.12	64.6	0.307	23.9	65.6	0.03	–22.1	69.1	0.286	–11.6
III	2-year dry	C (1963–1965)	0.270	86.5	0.127	–0.79	80.9	0.151	–3.97	84.8	0.140	8.19	87.2	0.127	–6.71	83.6	0.147	–14.5
	2-year wet	V (1954–1956)	0.542	63.1	0.365	3.31	74.3	0.305	–22.21	87.7	0.222	3.89	79.1	0.290	7.03	88.0	0.220	14.5
IV	10-year mixed	C (1953–1963)	0.448	73.5	0.241	1.84	69.1	0.259	–14.88	73.4	0.239	3.0	74.8	0.232	–0.3	63.4	0.281	–3.6
	11-year mixed	V (1963–1974)	0.381	70.8	0.268	18.43	60.5	0.311	–4.45	62.5	0.307	24.8	68.7	0.279	4.6	61.1	0.311	11.2

Table 3

Summary statistics of the calibration and verification runs for Pitman and Sacramento models applied to the Great Usuthu catchment of Swaziland. Both models were calibrated using SCE-UA and the HMLE objective function

Run no.	Data type	Test period (C = calibration, V = verification)	Mean flow (mm day^{-1})	Pitman model (PTM)			Sacramento model (SMA)		
				E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)
I	5-year wet	C (1953–1958)	0.505	76.7	0.234	−3.23	45.6	0.358	−33.39
	5-year dry	V (1963–1968)	0.326	76.1	0.221	18.02	45.0	0.330	−16.32
	11-year mixed	V (1963–1974)	0.381	67.2	0.284	17.04	36.1	0.397	−19.06
II	2-year wet	C (1954–1956)	0.542	67.5	0.343	−22.26	55.3	0.402	−31.29
	2-year dry	V (1963–1965)	0.270	62.8	0.211	−9.40	56.4	0.229	−11.93
	11-year mixed	V (1963–1974)	0.381	52.6	0.342	−10.64	37.8	0.392	−19.55
III	2-year dry	C (1963–1965)	0.270	53.0	0.238	−31.03	13.5	0.322	−48.11
	2-year wet	V (1954–1956)	0.542	60.1	0.380	−33.17	11.4	0.566	−51.57
IV	10-year mixed	C (1953–1963)	0.448	46.9	0.340	−24.23	28.8	0.394	−40.25
	11-year mixed	V (1963–1974)	0.381	66.0	0.290	−13.29	29.0	0.419	−32.42

6.1.4. SMAR model

Compared with the other models, SMAR (nine parameters) has a more restricted model structure. For example, even though it can operate up to five soil moisture zones, the capacity for each soil layer is set at 25 mm except for the lowest layer. Furthermore, it is only when all the soil layer zones are saturated will there be any runoff from the soil layers. This runoff, divided into surface runoff and groundwater by only one parameter G, has a ratio of surface runoff to groundwater that is more or less fixed by G. This relatively simple approach may work well with wet catchments but it is probably too restrictive for dry catchments, which experience a wide range of flow scenarios.

Since complex models can either do better or worse than a simple model such as SMAR, it seems that model complexity is less crucial than the model structure in modeling dry catchments (or wet catchments). On the whole, other than a few inappropriate features identified in this study, traditional CRR models with nine to 15 parameters are generally applicable to dry catchments if good calibration data are available, even though results from dry catchments are less predictable and are not expected to be as good as for temperate or wet catchments. In addition, to simulate high flows in medium sized, dry catchments accurately, it will be necessary to use data of time steps smaller than 1 day.

6.2. Comparison of objective functions

As mentioned in Section 6.1, Table 2 and Table 3 show that the DRMS objective function produced better statistics than the HMLE because DRMS places more weight on high flows while HMLE places more weight on low flows, and high flows have greater

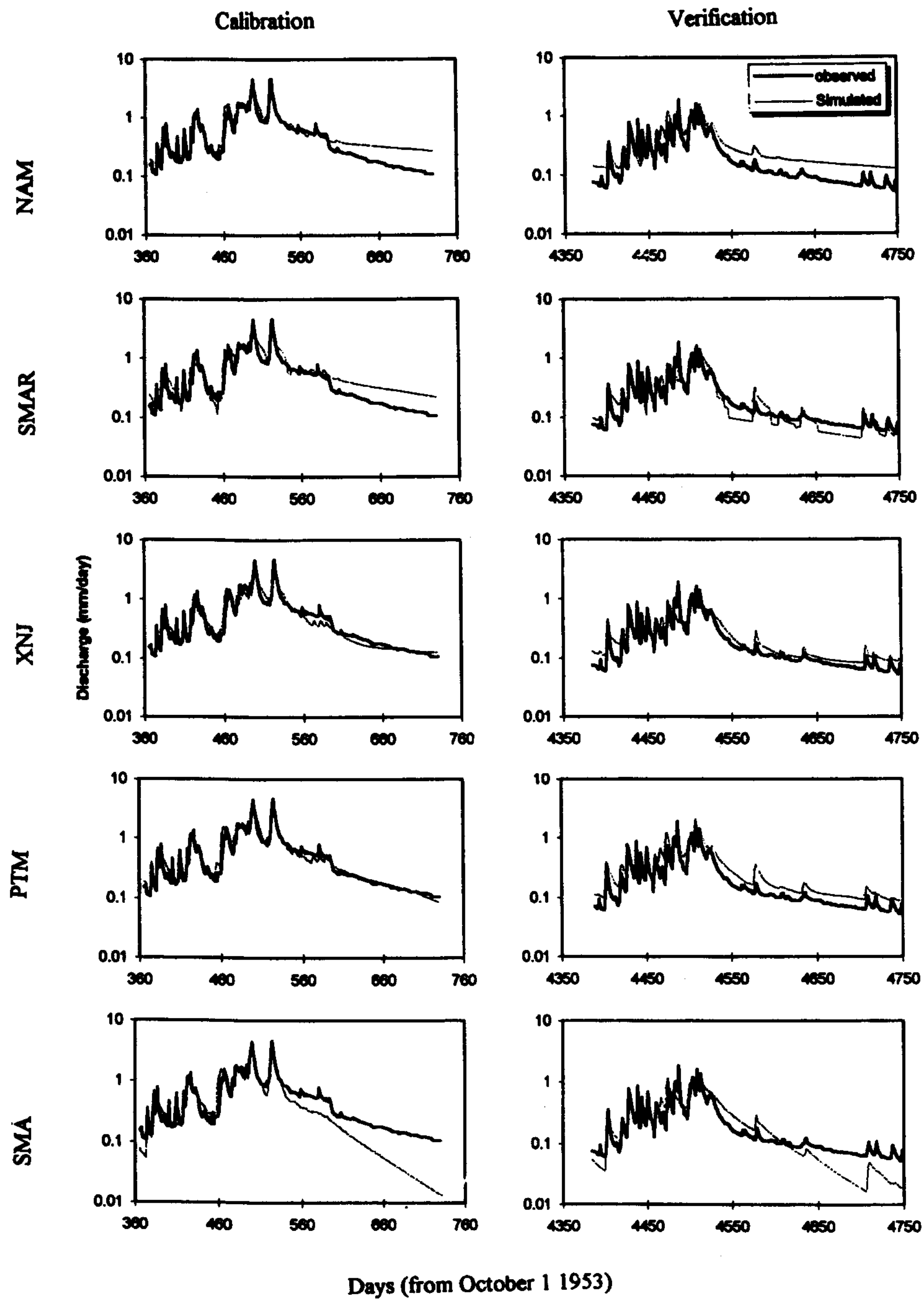

Fig. 3. Selected runoff hydrographs of the Test IV calibration and verification runs of the five models using DRMS as the objective function (which involves non-transformed data). Peak flows are generally accurately simulated but low flows are not.

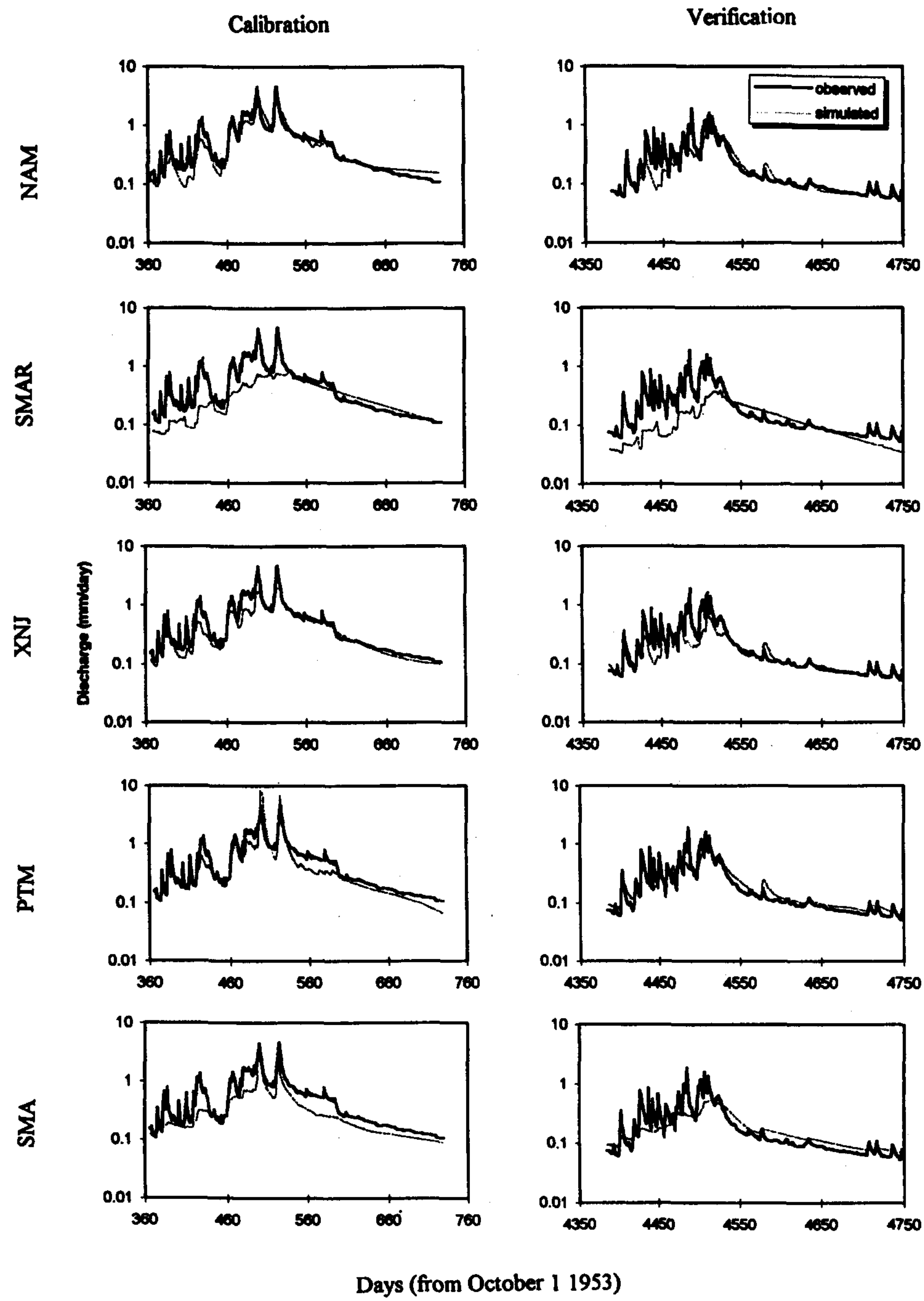

Fig. 4. Selected runoff hydrographs of the Test IV calibration and verification runs of the five models using HMLE as the objective function (which involves log-transformed data). Low flows are generally accurately simulated but high flows tend to be under-simulated.

Table 4

Summary statistics of the calibration and verification runs for Pitman, Sacramento, NAM, Xinanjiang, and SMAR models applied to the Ihimbu catchment of Tanzania. All the models were calibrated using SCE-UA, the DRMS, and the HMLE objective functions

Model type	Test period	DRMS			HMLE		
		E_f (%)	RMSE (mm)	Bias (%)	E_f (%)	RMSE (mm)	Bias (%)
SMAR	Calibration	86.79	0.173	−0.27	60.76	0.309	−20.73
	Verification	38.24	0.271	17.48	60.17	0.264	−12.00
XNJ	Calibration	89.72	0.155	−1.60	88.13	0.165	−3.61
	Verification	44.72	0.311	30.8	41.20	0.321	27.62
NAM	Calibration	89.76	0.155	−0.40	87.67	0.171	−3.21
	Verification	40.88	0.321	33.8	36.08	0.335	31.56
SMA	Calibration	85.22	0.184	−2.60	49.50	0.339	−24.6
	Verification	43.71	0.314	25.45	−20.56	0.460	8.7
PTM	Calibration	52.15	0.330	−7.16	26.08	0.411	−34.84
	Verification	42.93	0.316	−15.84	16.91	0.382	−36.55

impact on the objective functions. If low flows are more important, it may be justifiable using the HMLE. However, if high flows are desired, it would be wise to use an objective function that does not transform the flows, such as DRMS. It was found that HMLE is also computationally more expensive than DRMS, since it has an extra parameter to optimize (Sorooshian and Gupta, 1983). For some models it would be appropriate to carry out the calibration in two stages, using HMLE to calibrate low flow parameters, keep the low flow parameter values fixed, and calibrate the parameters responsible for high flows with DRMS. This approach was not attempted in this study.

Among models, optimization algorithms, data, and objective function, the objective function by far produced the largest difference in the result (compare the statistics given in Table 2 and Table 3). This is because non-transformed (DRMS) and log-transformed data (HLME) gave rise to substantially different response surfaces in the parameter search

Table 5

Summary statistics of the calibration and verification runs for Sacramento, NAM, Xinanjiang, and SMAR models applied to the Bird Creek of USA. All the models were calibrated using SCE-UA and the DRMS objective function

Model type	Test period	DRMS		
		E_f (%)	RMSE (mm)	Bias (%)
SMAR	Calibration	89.61	0.915	12.0
	Verification	71.99	0.532	20.0
XNJ	Calibration	90.66	0.868	24.4
	Verification	72.05	0.531	24.8
NAM	Calibration	81.90	1.204	15.5
	Verification	56.32	0.664	− 2.90
SMA	Calibration	92.20	0.775	1.31
	Verification	30.77	0.799	− 58.59

Table 6

Comparisons of PTM model parameters derived by SCE-UA for the Great Usuthu catchment using four sets of calibration data (listed as Tests I, II, III, and IV) and two objective functions, DRMS and HMLE

Test parameters	DRMS				HMLE			
	I	II	III	IV	I	II	III	IV
ST	396.6	354.3	162.1	354.3	372.8	398.5	405.9	344.5
FT	0.690	1.000	0.460	0.934	0.997	0.488	0.480	0.682
ZMINN	0.788	0.055	0.000	0.000	0.015	0.000	1.319	0.280
XMAXN	12.460	13.66	13.30	13.01	14.280	13.270	14.50	18.13
TL	8.269	3.074	4.169	4.957	3.714	10.930	5.53	4.33
LAG	1.634	1.279	1.169	0.644	1.427	1.853	1.33	1.741
GL	0.056	0.285	17.830	0.039	0.001	9.321	0.064	0.075
DIV	0.842	0.444	0.997	0.561	0.628	0.994	0.190	0.647
OBSQ	2.500	3.116	1.341	0.504	2.592	3.283	3.390	2.272
Calibration E_f (%)	65.8	89.1	86.5	73.5	76.7	67.5	53.0	46.9
Verification E_f (%)	64.6	81.1	63.1	70.8	76.1	62.8	60.1	66.0
	58.4	66.2			67.2	52.6		

space, causing some of the final calibrated parameters to be significantly different (compare PTM parameters obtained from DRMS and HMLE in Table 6).

6.3. *Effects of data variability on model calibration*

For the sake of illustration, parameter sets of PTM (Table 6) obtained for the Great Usuthu catchment using four sets of calibration data (Test I, II, III, and IV) and different objective functions are compared below. Except for ZMINN, which in a few instances took on the preset lower limits of 0.0, the eight parameter sets are clearly different from each other, and the degree of differences vary from parameter to parameter. This shows that calibrated parameters are data dependent. Under each objective function the choice of calibration data also gave rise to a fairly wide range of calibration results (as indicated by E_f and other statistics). None of the eight cases has an E_f above 90%, and under HMLE the E_f values are fairly low (between 47 to 77%). This shows that some calibrated parameters are more realistic than others but most likely none of the parameter sets is of the global optimum quality even though a global optimization algorithm, SCE-UA, was used in the calibration.

Ideally, if model parameters estimated are unique and realistic, the estimated parameters should be independent of the calibration data. In other words, if another calibration data set is used, the parameters estimated by the optimization method should be more or less the same (within the numerical accuracy and round-off errors). The concept of uniqueness used here is analogous to what Sorooshian and Gupta (1983), Sorooshian and Gupta (1985) referred to as "a model structure M parameterized by θ is *globally identifiable*, if and only if different parameter values of M give rise to different model output (streamflow)." In practice, because calibration involves adjusting the parameters until the difference between the simulated and observed streamflows is minimized, the final parameters are inevitably related to the calibration data. This data dependency feature is further complicated by the presence of insensitive parameters. If model parameters are insensitive or poorly identifiable, then different sets of parameters could essentially produce the same model output (Gan and Biftu, 1996). Sorooshian and Gupta (1985) attributed the parameter identifiability problems to model structure.

The calibrated models under Test II or 2 wet years (1954–1956) were verified twice, once with dry years and once with the common 11-year mixed data. For Test III or 2 dry years (1963–1965), the calibrated models were only verified once since there is an overlap between the calibration data period and the 11-year mixed data (1963–1974). Comparisons of Test II and Test III results (Table 2, Table 3, Table 6) indicate that to model the Great Usuthu catchment, wet years are preferred over dry years as calibration data because dry years may not contain enough high flows to sufficiently activate model parameters responsible for simulating high flows during model calibration. Wet years are more likely to contain both high and low flows and so they provide more ample information for model calibration. However, when a model calibrated with wet years was verified with dry years, the chances of getting an over-estimation of dry year flows tend to be higher than an under-estimation. Test II shows that moving from calibration to verification runs, there is an increase in BIAS in six out of nine cases. The reverse happened when verifying a dry-year calibration with wet years, e.g. Test III. In Test III, moving from calibration to verification runs, BIAS decreases in seven out of ten cases. For example, in Table 2 the BIAS of PTM calibrated with 2 wet years (Test II) was –0.55%, and it increased to 20.4% when verified with 2 dry years. When SMA calibrated with 2 dry years was verified with 2 wet years, the BIAS decreased from –3.97% to –22.2% (Test III). Since the use of wet years as calibration data tend to produce parameters that over-estimate streamflows at the verification stage or vice versa, this again shows the dependency of parameters on the data used for calibration.

6.4. *Effects of data length on model calibration*

Theoretically, a longer set of calibration data should achieve a better calibration because by going through a longer calibration experience, model parameters should be more accurately calibrated. Sorooshian et al. (1983) found this philosophy to be generally not true, as was shown by the results of Tests I, II, and IV, which used calibration data lengths of 2, 5, and 10 years, respectively, and which were tested with a common, 11-year verification run (1963–74). For example, of the 2 and 5 wet years and 10-mixed-year calibration cases, the PTM model produced an E_f of 66.2%, 58.4%, and 70.8% for the 11-year verification runs, respectively. The NAM had corresponding E_f values of 64.6%, 61%, and 62.5%, while XNJ's E_f values were 65.6, 64.2, and 68.7%. Therefore there is no indication that model performance is related to the calibration data length. In some instances, models calibrated with 2 years of data could perform better than models calibrated with 10 years of data. This shows that the data length is not that crucial, as long as it is not less than 1 hydrological year, and as long as the data used contain enough information for calibrating the parameters.

6.5. Comparisons of models and objective functions based on Ihimbu catchment

Due to a lack of data, the effects of data length and data variability was not carried out for the Ihimbu catchment of Tanzania. Under DRMS, all models except PTM show marginal differences in the calibration results for this catchment while under HMLE, XNJ, and NAM did better than other models (Table 4). However, the African model (PTM) did badly under both objective functions. Again, it shows that model performance is more closely related to the model structure and the objective function than model complexity since SMAR did better than PTM. For the Ihimbu catchment, the drop in model performance when moving from the calibration to the verification stage is mostly substantial. This means that either none of the parameter sets obtained is conceptually realistic or Ihimbu could have data problems. Since all five models did poorly at the verification stage, and with gross bias (either positive or negative bias), it is more likely that data problems cause the substantial drop in model performance. In a separate paper, Gan and Biftu (1996) showed that among eight catchments tested, the only trouble XNJ had was the verification stage for the Ihimbu catchment (E_f of 44.7% from SCE-UA). This reinforces our suspicion that Ihimbu has data problems.

For PTM, even the results at the calibration stage (particularly under HMLE) are also very poor. The trouble in getting adequate α and β values for converting daily to hourly rainfall (a feature of PTM) was the likely cause for PTM to perform more poorly than other models. This special requirement of PTM will render it inapplicable in areas where no hourly rainfall data is available. Despite the poor performance at the verification stage, XNJ seems to be doing better than other models, as was found in the Swaziland catchment.

6.6. Comparisons of models for Bird Creek, USA

From testing the Ihimbu catchment, we realized that without hourly rainfall data, the α and β parameters of PTM could not be properly determined and so only four models (SMA, NAM, XNJ, and SMAR) were applied to this third dry catchment. In addition, since it is already known that DRMS tends to produce better statistics than HMLE because it places more weight on the high flows, only DRMS was used. Results from the Bird Creek also show that it is feasible to apply traditional CRR models to dry catchments even though, on average, dry catchments are likely to perform poorer than wet catchments.

At the calibration stage, all four models achieved fairly high E_f statistics but at the verification stage the drop in the model performance was rather severe for NAM and SMA, especially SMA ($E_f = 30.8\%$). Since dry catchments undergo more variable hydrologic processes, certain models, because of their structural set-up, may not work well with dry catchments. In this case, Bird Creek is located in the US and yet the result of SMA turned out to be the worst even though, as the operational river forecast model of the US National Weather Service, SMA has been extensively used in the US. This is likely because SMA was mainly built for catchments in North America where the climatic conditions range from average to wet, where it is more feasible to rely only on one set of unitgraph ordinates. Apparently, this unconventional, one unitgraph approach seems problematic for dry catchments, irrespective of where the catchments are located.

7. Observations and summary

To study the hydrologic modeling of dry catchments, three such catchments, the Great Usuthu and the Ihimbu catchments located in Africa and the Bird Creek located in the US, were modeled with four or five conceptual rainfall–runoff models of different complexity and model structure. The observations are summarized below:

1. Albeit dry catchments generally undergo more complex and a wider range of hydrologic processes, it seems that good quality hydrologic data can still support the calibration of traditional CRR models with 10 to 20 parameters. On the whole, dry catchments are more sensitive to the model structure and harder to model than wet catchments. The model performance depends more on the model structure, the objective function used in automatic calibration, and data quality, than on model complexity (or number of parameters) or calibration data length. Also, it seems wet years provide better calibration data than dry years because the former contains more information (especially in terms of peak flows) than the latter.
2. XNJ has been consistently doing better than other models in all three catchments probably because it is the only model that considers the non-uniform distribution of runoff producing areas to simulate the runoff, which is crucial for dry catchments. Complex models such as SMA and simple models such as SMAR did relatively poorly because SMA only uses one set of unitgraph (unconventional unit hydrograph) ordinates to route low and high flows together, and SMAR may have a model structure that is a little too simple for dry catchments. The need for hourly rainfall data to determine two of its model parameters makes PTM inapplicable in places without hourly data.
3. The objective function is an important element in calibrating CRR models automatically. A simple least-squares objective function, DRMS (which puts more weight on the high flows) is preferred over HMLE (which puts more weight on the low flows), if an accurate simulation of high flows is more important than low flows, and vice versa.

Strictly speaking, no known scientific theory has been established to relate the rainfall–runoff transformation process at a catchment scale and so CRR models have been widely used to model this process. These models attempt to extend known theories and observations of atmospheric dynamics, fluid mechanics, and soil physics at point scales, to hydrologic processes at catchment scales. To some researchers such as Dooge (1988), this may not be the approach to pursue. In modeling dry catchments, we recommend that users be careful in choosing the model(s), the objective function(s), and calibration data to be applied to their catchments. As shown in this study, even a model that has been extensively applied to many catchments, such as SMA, may not work well under dry climatic conditions.

Acknowledgements

This work was partially supported by the Natural Science and Engineering Research Council of Canada. The second author was also supported through an exchange program between the Canadian International Development Agency (CIDA) and Swaziland.

Drs S. Sorooshian, Q. Duan, and V.K. Gupta of Arizona, USA, kindly provided the computer software for the optimization algorithm SCE-UA. Dr W.V. Pitman of the University of Witwatersrand provided the computer codes of the Pitman model. The review comments of Dr. Ian Cordery and an anonymous reviewer have improved the quality of this paper. The test data for the catchment in Swaziland were obtained from the Computing Center for Water Research of South Africa, while the test data for the catchments in Tanzania and the US were obtained from the Hydrology Department, University College of Galway, Ireland.

References

Beven, K., 1989. Changing ideas in hydrology – the case of physically based models. J. Hydrol., 10: 157–172.

Brazil, L.E., 1980. Unit Hydrograph. National Weather Service River Forecast System, User's Manual, II.4-UNIT-HG-1.

Burnash, R.J., Ferral, R.L. and Mcguire, R.A., 1973. A Generalized Stream Flow Simulation System, Conceptual Modeling for Digital Computers. US NWS, Sacramento, CA.

Danish Hydraulic Institute, 1982. NAM Model Documentation. 82-892, JCR/Skn.

Dooge, J.C.I., 1988. Modeling the behaviour of water. In: E.R.C. Reynolds and F.B. Thompson (Editors), Forests, Climate, and Hydrology: Regional Impacts. United Nations University.

Duan, Q., Sorooshian, S. and Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall runoff models. Water Resour. Res., 28(4): 1015–1031.

Gan, T.Y., 1988. Application of scientific modeling of hydrological response from hypothetical small catchments to assess a complex conceptual rainfall runoff model. Water Resources Series Tech. Rep. No. 111, Dept. of Civil Eng., Univ. Washington, Seattle.

Gan, T.Y. and Biftu, G.B., 1996. Automatic calibration of conceptual rainfall–runoff models: optimization algorithms, catchment conditions, and model structure. Water Resour. Res., 32(12): 3513-3524.

Gan, T.Y. and Burges, S.J., 1990. An assessment of a conceptual rainfall–runoff model's ability to represent the dynamics of small hypothetical catchments, 1, Models, model properties, and experimental design. Water Resour. Res., 26(7): 1595–1604.

Gupta, V.K. and Sorooshian, S., 1985. The automatic calibration of conceptual models using deterministic based optimization algorithms. Water Resour. Res., 21(4): 473–485.

Hendrickson, J., Sorooshian, S. and Brazil, L.E., 1988. Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall–runoff models. Water Resour. Res., 24(5): 691–700.

Hewlett, J.D., 1982. Principles of Forest Hydrology. University of Georgia Press, Athens.

Holland, S.H., 1975. Adaptation in Natural and Artificial Systems. Univ. Michigan, Ann Arbor.

Hornberger, G.M., Beven, K.J., Cosby, B.J. and Sappington, D.E., 1985. Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. Water Resour. Res., 21: 1841–1859.

Hughes, D.A. and Herald, J., 1987. The application of deterministic catchment hydrological model in contemporary problems and suggestions for a more unified approach. In: Proc. 1987 Hydrol. Science Symposium, September 1987, Rhodes University, Grahamstown. Vol. II, Rhodes University, Grahamstown, pp. 500–508.

Jakeman, A.J. and Hornberger, G.M., 1993. How much complexity is warranted in a rainfall–runoff model? Water Resour. Res., 29(8): 2637–2649.

Kachroo, R.K., 1992. River flow forecasting, Part 5, Applications of a conceptual model. J. Hydrol., 133: 141–178.

Loague, K.M. and Freeze, R.A., 1985. A comparison of rainfall–runoff modeling techniques on small upland catchment. Water Resour. Res., 21: 229–248.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models; Part 1-A discussion of principles. J. Hydrol., 10(3): 282–290.

Nelder, J.A. and Mead, R., 1965. A simplex method for functional minimization. Comput. J., 9: 308–313.

O'Connell, P.E., Nash, J.E. and Farrell, J.P., 1970. River flow forecasting through conceptual models, Part 2, The Brosna catchment at Ferbane. J. Hydrol., 10: 317–329.

Pitman, W.V., 1976. A mathematical model for generating daily stream flow from meteorological data in South Africa. Rep. 2/76, HRU, Univ. Witwatersrand, Johannesburg.

Pitman, W.V., Middleton, B.J. and Midgley, D.C., 1981. Surface water resources of South Africa. Hydrological Research Unit Rep. No. 9/81, Vol. IV, Department of Civil Engineering, University of Witwatersrand, South Africa.

Sorooshian, S. and Gupta, V.K., 1983. Automatic calibration of conceptual rainfall–runoff models: The question of parameter observability and uniqueness. Water Resour. Res., 19(1): 260–268.

Sorooshian, S. and Gupta, V.K., 1985. The analysis of structural identifiability: Theory and application to conceptual rainfall–runoff models. Water Resour. Res., 21(4): 487–495.

Sorooshian, S., Gupta, V.K. and Fulton, J.L., 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall runoff models – influence of calibration data variability and length on the model credibility. Water Resour. Res., 19(1): 251–259.

Sorooshian, S., Duan, Q. and Gupta, V.K., 1993. Calibration of rainfall–runoff models: Application of global optimization to the Sacramento soil moisture accounting model. Water Resources Res., 29(4): 1185–1194.

Zhao, R.J., Zhuang, Y.L., Fang, L.R., Liu, X.R. and Zhang, Q.S., 1980. The Xinanjiang model. In: Hydrological Forecasting, Proceedings of the Oxford Symposium, April 1980. IAHS Publ. No. 129.