WATER RESOURCES RESEARCH, VOL. 37, NO. 9, PAGES 2345–2352, SEPTEMBER 2001

A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications

Teemu S. Kokkonen
Laboratory of Water Resources, Helsinki University of Technology, Helsinki, Finland

Anthony J. Jakeman
Centre for Resource and Environmental Studies, Australian National University, Canberra, ACT, Australia

Abstract. The aim of the present paper is to compare metric and conceptual approaches to rainfall-runoff modeling in terms of calibration and simulation performances and parameter invariance. This is investigated by applying two models of equal complexity (i.e., possessing the same number of parameters), but with different levels of "conceptualization," to two catchments with different climatology. Level of conceptualization is understood as the degree to which the model structure and its parameters can be related to catchment-scale hydrological processes. The results suggest that the model with less conceptualization provides, in general, a more accurate reproduction of streamflow, even on independent data sets, but this difference only becomes clear when models are applied to the drier catchment. The paper corroborates that the more process complexity one wants to include in the model structure, the more types of data and higher information content are required to estimate the process parameters and to test the model performance. When only rainfall-runoff data are available, it is difficult to justify substantial conceptualization of complex processes.

1. Introduction

1.1. Model Classification

The simulation of rainfall-runoff relationships has been a prime focus of hydrological research for several decades and has resulted in an abundance of models having been proposed. Following *Beck* [1991], these models can crudely be classified as metric, conceptual, and physics-based.

Metric models are strongly observation-oriented seeking to characterize system response by extracting information from the existing data. They are constructed with little or no consideration of the features and processes of the hydrological system. The development of unit hydrograph theory [*Sherman*, 1932] is the foundation of metric rainfall-runoff models. Unit hydrograph theory is based on the linearity assumption between rainfall excess and streamflow. This relationship can be represented by the continuous time convolution of rainfall excess with an instantaneous unit hydrograph to give continuous flow. Often this flow is thought to be only the storm flow, or direct flow, component, and hence some representation for base flow is added to attain the total streamflow. Techniques for the separation of different flow components have been, albeit intuitively reasonable, inherently arbitrary [*Littlewood and Jakeman*, 1994]. For example, base flow can be derived by passing the streamflow through a low-pass numerical filter. *Natale and Todini* [1977] used two parallel linear response functions with a switch between the two, dependent on aggregated antecedent precipitation. More recently, sophisticated algorithms such as the Kalman filter [*Beck et al*., 1990] and instrumental variable methods [*Jakeman et al*., 1990] have been applied to conceptual stores to perform the separation. One of the major strengths of methods relying on the unit hydrograph is their minimal data requirements.

Conceptual models describe all of the component hydrological processes perceived to be of importance as simplified conceptualizations. This usually leads to a system of interconnected stores, which are recharged and depleted by appropriate component processes of the hydrological cycle. In the last 30 years, hundreds of alternative conceptual models have been developed, including the Stanford Watershed Model [*Crawford and Linsley*, 1966], the Tank model [*Sugawara et al*., 1983], the *Boughton* [1984] model, MODHYDROLOG [*Chiew and McMahon*, 1994], and Hydrologiska Bryäns Vattenbalansavdelning [*Bergström*, 1995]. The more component processes that are included in the model the higher the risk of overparameterization. The associated effects of parametric uncertainty in environmental modeling are extensively documented in the literature [*Freer et al*., 1996; *Johnston and Pilgrim*, 1976; *Spear et al*., 1994].

Finally, physics-based models attempt to mimic the hydrological behavior of a catchment by using the concepts of classical continuum mechanics. The governing partial differential equations are normally solved numerically by applying finite difference or finite element computational schemes. *Freeze* [1972] developed the first such model, in which finite difference methods were used to solve Richards' equation for unsaturated flow in two dimensions to represent hillslope processes. More recently, models such as the Institute of Hydrology Distributed Model [*Beven et al*., 1987] and the Système Hydrologique Européen model [*Abbott et al*., 1986; *Bathurst*, 1986] have been developed with essentially similar mathematical formulations. Physics-based models are appealing to some, since they provide a mathematically idealized representation of the real phenomenon. At the same time, they require massive amounts of data which are difficult to obtain and have high computational demands. *Beven* [1989], *Binley*

Paper number 2001WR000299.
0043-1397/01/2001WR000299$09.00

and Beven [1989], and *Grayson et al.* [1992] discuss the applicability of physics-based models.

1.2. Use of Rainfall-Runoff Models

Understanding runoff generation has a significant role in catchment hydrology. Some of the tasks envisioned for rainfall-runoff models are of a purely hydrological nature, such as real-time flood forecasting, design flood estimation, and assessment of the reliability of natural water resources. However, increasingly, outputs of hydrological models are used to investigate wider environmental problems. These include water quality issues in surface and groundwaters [*Christophersen and Wright*, 1981; *Cosby et al.*, 1985], ecological studies, and providing boundary conditions for models dealing with atmospheric general circulation [*Dumenil and Todini*, 1992; *Eagleson*, 1986; *Kuhl and Miller*, 1992; *Leung et al.*, 1996; *Wood et al.*, 1992].

All these tasks set different requirements on models. In flood forecasting, event response models are often sufficient, thus eliminating complexities of continuous moisture accounting, evaporation, and long-term dynamics. Water quality studies frequently require explicit knowledge of different flow pathways [e.g., *Chapman et al.*, 1993], so models used for such investigations must be capable of making the distinction between direct runoff, subsurface runoff, and groundwater flow. When hydrological models are linked with general circulation models, evaporation must be explicitly represented.

Selecting a model with an appropriate level of model complexity for a particular problem is far from straightforward. An increase in model complexity does not only mean an inevitable increase in data requirements and computational costs, but it also easily results in ill conditioning and nonidentifiable parameters [*Dietrich et al.*, 1993]. Nevertheless, hydrologists are constantly faced with problems where more detailed knowledge and quantification of the component processes of the hydrological cycle are essential.

1.3. Objective of this Study

In this paper the performance of two modeling approaches is discussed. Both are variations of a model called identification of hydrographs and components from rainfall, evaporation and streamflow data (IHACRES) [*Evans and Jakeman*, 1998; *Jakeman and Hornberger*, 1993]. One is a more conceptual approach which splits rainfall into evapotranspirational losses and streamflow, while the other one is of a more metric nature with no ambition to quantify evapotranspiration in an explicit manner. The level of complexity (defined herein as the number of parameters) is the same in both models.

Calibration and simulation performance of these models is investigated using data from two climatologically different catchments. One is a relatively humid catchment in North Carolina in the United States, and the other catchment is a low-yielding ephemeral catchment in Western Australia.

The consistency of parameter estimates yielded by the models is also addressed. Ideally, parameters should be independent of climate sequence in model calibration periods [*Gan and Burges*, 1990]. To investigate the extent to which this is true for the two models under study, both models are calibrated on 45 three-year time periods using the data from the North Carolina catchment, and the variability of the parameter estimates obtained from different calibration periods is assessed.

The aim of the study is to compare the performance aspects of two models of equal complexity but with different levels of conceptualization. The more conceptual model can deliver as a by-product an estimate of the catchment evapotranspiration, while the metric model merely attempts to predict the time series of streamflow.

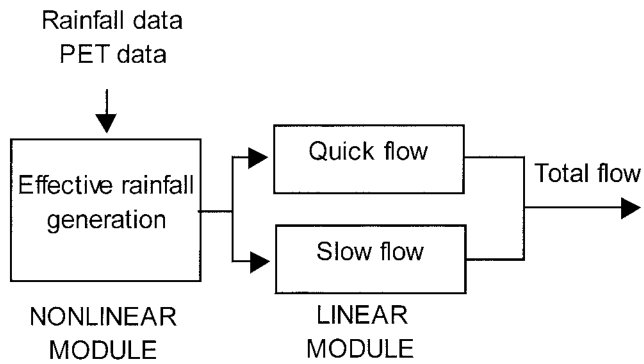

Figure 1. Systems diagram of the rainfall-runoff model.

2. Models

Both models can be divided into nonlinear and linear modules. The nonlinear or rainfall loss module converts rainfall to rainfall excess or effective rainfall, which is defined as the share of rainfall that eventually becomes streamflow. The linear module represents the transformation of rainfall excess to streamflow.

The structure of the linear module is the same in both models. IHACRES allows flexible configuration of linear stores connected in parallel and/or series. Using statistical identification procedures, the most appropriate form for the North Carolina catchment is two stores in parallel [*Jakeman and Hornberger*, 1993]. *Ye et al.* [1997] found that one linear store is sufficient for the ephemeral Western Australian catchment. The equations underlying the linear module, and its link to a transfer function model, are given by *Jakeman et al.* [1990] and *Jakeman and Hornberger* [1993]. Figure 1 shows a systems diagram for the model with two stores in parallel.

The models differ in the structure of the nonlinear loss module. The more conceptual loss module [*Evans and Jakeman*, 1998] is basically a store which is recharged by precipitation P and depleted by evapotranspiration E and effective rainfall U. However, the state of the store is not expressed as water level but as a catchment moisture deficit CMD. The catchment moisture store accounting scheme is given at time step k by

$$\mathrm{CMD}_k = \mathrm{CMD}_{k-1} - P_k + E_k + U_k. \qquad (1)$$

Evapotranspiration at time step k, E_k, is characterized here as a function of potential evaporation (or temperature, T_k, as a surrogate for potential evapotranspiration, PET_k) and catchment moisture deficit. Effective rainfall is assumed to be dependent on catchment moisture deficit only. The parameterizations for evapotranspiration and effective rainfall are defined as

$$E_k = c_1 \, \mathrm{PET}_k \exp\left(-c_2 \, \mathrm{CMD}_k\right) \qquad (2)$$

$$U_k = \begin{cases} c_3 - \mathrm{CMD}_k & \mathrm{CMD}_k < 0 \\ \dfrac{-c_3}{c_4} \mathrm{CMD}_k + c_3 & 0 < \mathrm{CMD}_k < c_4, \\ 0 & \mathrm{CMD}_k \geq c_4 \end{cases} \qquad (3)$$

where c_1, c_2, c_3, and c_4 are model parameters.

In the statistical loss module the effective rainfall is computed using

$$U_k = s_k^p P_k, \tag{4}$$

where p is a model parameter which modulates the effect of the catchment wetness index s_k. The index s_k is calculated by a weighting of the precipitation time series, the weights decaying exponentially backward in time:

$$s_k = cP_k + (1 - \tau_w^{-1})s_{k-1}. \tag{5}$$

The parameter τ_w is approximately the time constant or, inversely, the rate at which the catchment wetness declines in the absence of rainfall. The parameter c represents the increase in storage index per unit rainfall in the absence of evapotranspiration. It is chosen so that the volume of effective rainfall is equal to the total streamflow volume over the calibration period. To account for fluctuations in evapotranspiration, the following simple function of potential evaporation is used:

$$\tau_w(\mathrm{PET}_k) = \tau_w \exp\left[(c_b - \mathrm{PET}_k) f\right], \tag{6}$$

where f is a PET (or temperature) modulation parameter on the rate of evapotranspiration and c_b is an arbitrary constant at which value $\tau_w(c_b) = \tau_w$.

From now on, for the sake of simplicity, the model of a more conceptual nature is called the "conceptual model," whereas the more statistical model is referred to as the "metric model." This characterization is a relative rather than an absolute one.

3. Catchments and Data

The two catchments selected for the model comparison are Coweeta 36 in North Carolina, United States, and Salmon Creek in Western Australia. They are about the same size (<1 km^2) but are climatologically very different.

Coweeta 36 (0.49 km^2) is one of the experimental catchments in the Coweeta Hydrological Laboratory. It is a high-elevation, steeply sloping catchment with shallow soils and a high annual yield (mean annual yield 0.75). The mean annual precipitation is 2220 mm [*Swift et al.*, 1988]. Coweeta 36 is a control catchment, and thus it has undergone little land use change. The catchment is covered by hardwood forest.

Salmon Brook (0.82 km^2) is located in the low-relief Darling Range of southwestern Western Australia, and it is a Western Australian Benchmark Catchment. The mean annual rainfall is 1200 mm. The region is dominated by jarrah (Eucalyptus marginata) forest. Surface soils are predominantly highly permeable sands and gravels [*Ye et al.*, 1997]. The catchment has suffered minimal land cover interference over the period of record analyzed.

In order to assess the calibration and simulation performance of the two models, two periods were selected for each catchment. In Coweeta the first was from October 4, 1947, to October 17, 1950, and the second was from September 20, 1984, to November 1, 1987. The first period is a wet period having an average rainfall of 7.2 mm/d, which is well above the average over the entire record (6.0 mm/d). It also includes the largest daily runoff value in the whole record (115 mm). The second period is relatively dry. The average rainfall is 4.6 mm/d, and the largest daily runoff of this period is half of that of the first period, 57.5 mm.

For Salmon Brook we used the same two 5-year calibration periods as *Ye et al.* [1997] to enable a direct comparison with the results of that study. The two periods of record analyzed were from January 1, 1974, to December 31, 1978, and from January 1, 1979, to December 31, 1983. The period of 1974–1983 is an interesting period for the region climatologically. The years 1973 and 1974 had above-average rainfalls, and the next 4 years consistently had below-average rainfalls. The subsequent 5 years experienced above-average rainfalls again. In other words, 1974–1978 represents a dry period, and 1979–1983 represents a wet period. To assess the consistency of parameter estimates of the models, 47 years of Coweeta data (1944–1991) were divided into 45 three-year periods, each of which overlapped the adjacent periods by 2 years.

4. Assessment Criteria

The deviations of observed flow q_i from modeled flow $\hat{q}_i$ were assessed using three statistics. These were bias (B), efficiency (E), and the O statistic, defined as

$$B = N^{-1} \sum_i (q_i - \hat{q}_i) = N^{-1} \sum_i e_i, \tag{7}$$

$$E = 1 - \sigma_e^2/\sigma_q^2, \tag{8}$$

$$O = N^{-1} \sum_i (\sqrt{q_i} - \sqrt{\hat{q}_i})^2. \tag{9}$$

In the above, N is the number of observations in the time series. The term σ_e^2 is the variance of the residuals e_i between the observed and modeled flow, and the term σ_q^2 is the variance of the observed discharge.

These three statistics judge different aspects of model performance. Bias is indicative of a model's ability to predict the volume of stream discharge. Lower values of bias are preferred for predicting water supply to reservoirs, for example. Efficiency indicates a model's ability to explain the variance of observed streamflow and is biased toward high flows, penalizing larger absolute discrepancies more than lower ones. The O statistic [e.g., *Chapman*, 1970] involves a compromise between fitting large and small runoff events by using a square root function to equilibrate different sizes of discharge.

5. Results

Both models were calibrated at a daily time step for the two catchments. In Coweeta the models were calibrated on two 3-year sequences of data, and in Salmon Brook the calibration periods were 5 years long. IHACRES is often calibrated using just 2 years of daily data. To assess the simulation performance, the models calibrated on one set of data of each catchment were then applied to the other one of the same catchment.

The calibration results are summarized in Table 1. For Coweeta, no substantial difference in calibration performance between the two models was detected in terms of efficiency and O value. In the first calibration period the metric model was marginally superior in efficiency (86% versus 85%), while in the other period it was outperformed by the conceptual model (79% versus 81%). In both calibration periods the metric model yielded lower absolute bias and O values. In the ephemeral Salmon catchment the metric model was clearly better in terms of efficiencies, 82% and 87% compared with 76 and 82% for the conceptual model. The absolute values of bias for the

Table 1. Calibration Results

	Coweeta					
	First Period (1947–1950)			Second Period (1984–1987)		
	Efficiency, %	Bias, mm	O, mm	Efficiency, %	Bias, mm	O, mm
Metric	86	0.03	0.13	79	0.08	0.10
Conceptual	85	−0.14	0.16	81	−0.15	0.11
	Salmon Brook					
	First Period (1974–1978)			Second Period (1979–1983)		
	Efficiency, %	Bias, mm	O, mm	Efficiency, %	Bias, mm	O, mm
Metric	87	0.00	0.05	82	−0.02	0.04
Conceptual	82	0.10	0.08	76	0.10	0.09

metric model are lower, ranging from 0.00 to 0.02, whereas the bias for the conceptual model is 0.10 for both time periods. The metric model also yields better O values (0.04–0.05) than the conceptual model (0.08–0.09).

Simulation results are shown in Table 2. In simulation mode the metric model outperforms the conceptual one in all cases (i.e., both catchments and both periods) when looking at the efficiency and in three out of four cases when bias and O values are considered. In Salmon Brook the advantage in favor of the metric model is more distinct than in the humid Coweeta catchment.

In addressing the consistency of parameter estimates, only the parameters of the linear module were used (Table 3). The linear module is conceptually the same in both models facilitating the comparison of results. The Coweeta data were used for this study because of its longer record length. In Coweeta the linear module was configured to consist of two parallel linear stores, which are fully described through three parameters [*Jakeman et al.*, 1990]: time constant of the quicker store τ_q, time constant of the slower store τ_s, and relative volumetric throughput of the slower store ν_s (relative volumetric throughput of the quicker store $\nu_q = 1 - \nu_s$). The metric model consistently yielded larger time constants and a higher ratio of "slow flow." The parameter variability through time is assessed by looking at the coefficient of variation (CV). The CV of the "quick flow" time constant τ_q of the conceptual model is 0.32, which is twice the value of that of the metric model (0.16). The conceptual model also has considerably higher variability in the volumetric throughput of slow flow ν_s, the CV being 0.19, while for the metric model it is 0.11. However, the conceptual model yields a less variable slow flow time constant τ_s (CV of 0.29) than the metric model does (CV of 0.39). The average values, standard deviations (SD), and coefficients of variation of all the parameters of the linear module are listed in Table 3.

6. Discussion of Results

In the humid Coweeta catchment neither of the models is clearly superior to the other one. Although the metric model often, but not always, yields higher efficiencies and lower bias and O values, the difference between the performance statistics of these two models is relatively smaller than for the Salmon Brook catchment. In Salmon Brook it is clear that the metric model outperforms the conceptual one where all the calibration and simulation statistics of the metric model are superior, sometimes quite considerably. These results suggest that the metric model in general provides a more accurate prediction of the streamflow. However, this difference only becomes appreciable in the ephemeral Salmon Brook catchment which represents a much more demanding modeling task than the Coweeta catchment where base flow is relatively large, total flow is less variable, and soil moisture accounting is less complex. In the Salmon Brook flow record of 10 years used in this study, the share of zero flow days is 52%, while 91% of daily runoffs are <1 mm. Thus the information content of the data is small which sets a greater challenge for the estimation

Table 2. Simulation Results

	Coweeta					
	First Period (1947–1950)			Second Period (1984–1987)		
	Efficiency, %	Bias, mm	O, mm	Efficiency, %	Bias, mm	O, mm
Metric	83	−0.60	0.16	76	0.30	0.12
Conceptual	81	−0.01	0.15	71	−0.49	0.20
	Salmon Brook					
	First Period (1974–1978)			Second Period (1979–1983)		
	Efficiency, %	Bias, mm	O, mm	Efficiency, %	Bias, mm	O, mm
Metric	85	0.01	0.04	77	−0.03	0.05
Conceptual	77	0.10	0.09	52	0.11	0.10

Table 3. Statistics of Parameter Variability: Mean, Standard Deviation, and Coefficient of Variation[a]

	τ_q, days			τ_s, days			ν_s		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
Metric	2.11	0.33	0.16	45.06	17.73	0.39	0.52	0.05	0.11
Conceptual	1.49	0.48	0.32	26.11	7.63	0.29	0.38	0.07	0.19

[a]Abbreviations are as follows: τ_q, time constant of the quicker store; τ_s, time constant of the slower store; ν_s, volumetric throughput of the slower store; SD, standard deviation; and CV, coefficient of variation.

of parameters. As pointed out by *Sorooshian et al.* [1983], rather than length, it is the quality of information in the data which is important.

It is worth noting that in Salmon Brook there were no data available to "run in" the soil moisture index (metric model) and the catchment moisture deficit (conceptual model) when models were applied to the first 5-year period. Both soil moisture index and catchment moisture deficit were initialized to be zero, which probably gives some advantage to the metric model. A zero value of soil moisture index corresponds to a dry catchment, whereas when catchment moisture deficit is zero the catchment is wet. It is very likely that at the beginning of the record the catchment was quite dry. The record starts January 1, and in that location, 80% of the annual rain falls between May and October. To investigate how much of an impact this initialization might have had on the results, both models calibrated on the second 5-year period were used to simulate the last 4 years of the first 5-year sequence. In this situation a whole year of data could be used to estimate the initial catchment moisture deficit, both in calibration and simulation. The metric model uses 100 days to start the soil moisture index. In this simulation the metric model yields an efficiency of 73%, while that yielded by the conceptual model is 27%. This difference is much greater than when simulating over the entire first 5-year period 1974–1978 (see Table 2). The year 1974 had above-average rainfalls, and the subsequent 4 years had below-average rainfalls. So even when the catchment moisture deficit is initialized more realistically, it is evident that the metric model predicts the streamflow more accurately in the low-yielding Salmon Brook catchment. Omitting the wet year 1974 from the simulation makes the difference more distinct.

The models investigated in this paper are fairly simple in having only six parameters each. The conceptual model has more ambitious objectives since it delivers not only an estimate of the streamflow but also an estimate of evapotranspiration. Therefore it may have to compromise in its accuracy of streamflow prediction, when compared to the metric model, which possesses the same level of complexity but has only been designed to predict streamflow. This assumption is supported by *Ye et al.* [1997], who found the metric model of this study to be very competitive when compared with more complex conceptual models consisting of interlinked stores. In the nonlinear module of the metric model the soil moisture index is exponentiated, and mass balance is forced (only in calibration mode) by scaling the sum of the effective rainfall with the aid of the parameter c to be equal to the total streamflow of the calibration period. None of these operations can be given any deep conceptual meaning, but they do allow the effective rainfall time series to be influenced in a plausible and simple nonlinear way by antecedent conditions.

In all 45 calibration periods which were used to assess the parametric invariance of the models, the metric model yields longer time constants and a larger ratio of slow flow. This is explained by the delay which the conceptual model has in its rainfall excess generation. Since the metric model has no delay in the nonlinear module, its transformation of the rainfall excess into streamflow has to be slower. Particularly when rainfall-runoff models are used to identify changes of hydrological behavior in a catchment, that is, when a shift in the value of an individual parameter is interpreted to reflect a change in the hydrological response, the robustness of the parameter estimates becomes an essential property. In two out of three parameters investigated, the metric model had less variability in the calibrated value of the parameter, and in one parameter the conceptual model delivered more consistently calibrated values.

Edijatno et al. [1999] discuss a model which they call an "empirical" model. An empirical model is not based on physical or conceptual reasoning, but it is only a combination of mathematical operators which have been selected through comprehensive testing in a large array of catchments. Their model has only three adjustable parameters, and still it achieved worthwhile results on a set of 140 catchments in France. The results of the present study also suggest that when the objective is merely to predict streamflow, the use of a simple model with only a few parameters, not paying attention to the conceptual interpretation of the model, could result in a more accurate reproduction of streamflow. However, in some problems an explicit representation of the component hydrological processes is required. For example, the conceptual model of this paper delivers an estimate of the amount of evapotranspiration in the catchment, which the metric model cannot do. So choice of a model always has to be based on the problem at hand.

Finally, problems can be encountered if the conceptual model is used for simulating evapotranspiration when it has been calibrated against flow data only. For demonstrational purposes, consider replacing (2) of the conceptual model by an equation where the PET is risen to the fourth power, i.e.,

$$E_k = c_1 \,\mathrm{PET}_k^4 \exp(-c_2 \,\mathrm{CMD}_k). \tag{10}$$

This model was calibrated on the Coweeta catchment for the first (1947–1950) of the 3-year periods used to assess the performance of the models. In this modified form, slightly better calibration statistics were obtained, the efficiency being 87% versus 85% of the original model. The difference in the modeled streamflow between these two models is very small (see Figure 2). However, they generate drastically different evapotranspiration time series. In Figure 3 the computed evapotranspiration time series of the two models are shown. It is evident that, when compared to the original model, the modified model gives substantially higher values for evapotranspiration in summer and gives much lower values in winter. The

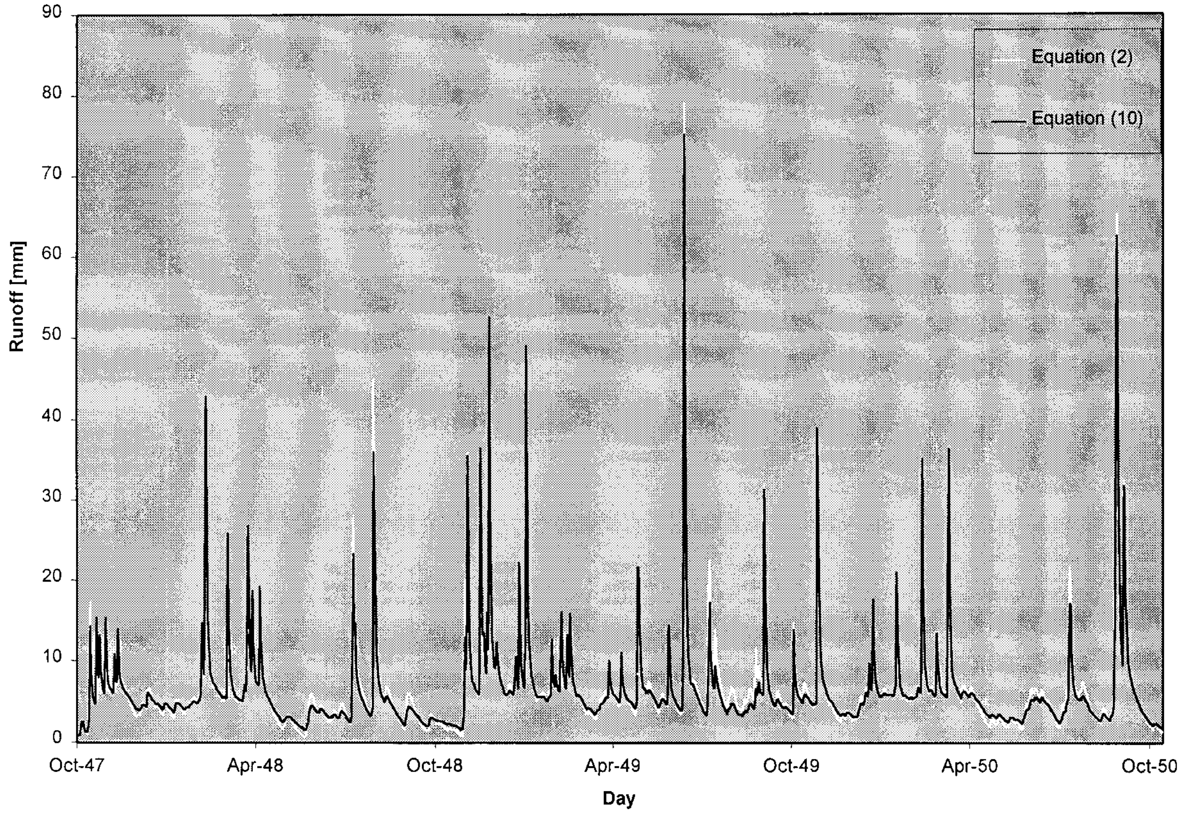

Figure 2. Comparison of runoff time series in Coweeta between two different evapotranspiration representations.

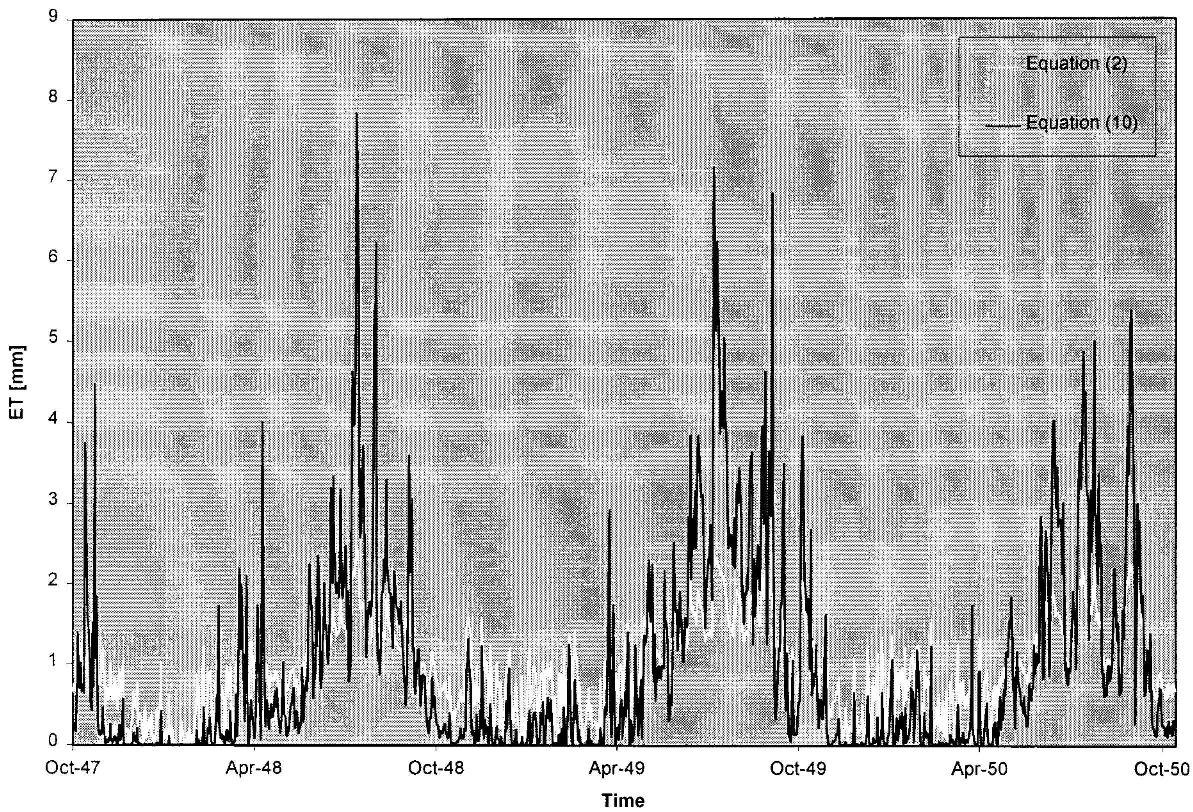

Figure 3. Comparison of evapotranspiration time series in Coweeta between two different evapotranspiration representations.

aim of this exercise was to demonstrate that calibration against flow data only does not help in assessing the structural adequacy of the representation of evapotranspiration. Large differences between models in terms of evapotranspiration prediction can occur, while streamflow predictions are not substantially different.

To avoid accepting a model which simulates evapotranspiration poorly, evapotranspiration data, or other constraining information, should be used wherever possible in the calibration process. This also applies to the representation of other catchment processes, such as differentiating between different water pathways (e.g., surface runoff, interflow, and groundwater flow). The need for multiobjective optimization of hydrologic models has been recognized [*Gupta et al*., 1998; *Yapo et al*., 1998]. Unfortunately, the availability of measured data for estimating the parameters in such processes is generally poor. One possible starting point is to make use of the collections of deep weighing lysimeter data which have been gathered for agricultural rather than hydrological purposes. *Chapman and Malone* [1999] used such data to compare the quality of groundwater recharge simulations of 11 models.

7. Conclusions

In this paper we have considered the performance of two rainfall-runoff models possessing a different levels of conceptualization. Conceptualization is understood as the degree to which the model structure and its parameters can be related to the catchment-scale hydrological processes. Increasing the level of conceptualization enables the model to take into account various fluxes of the hydrological cycle, whose explicit representation is frequently required in hydrology-related environmental problems.

Both models are fairly simple, having only six parameters. The results presented here indicate, first, that if the only objective is to reproduce the streamflow, incorporating more conceptualization into a simple rainfall-runoff model may result in decline of the model performance. The less conceptual model of this study provided generally more accurate representation of streamflow. However, the superior performance was most apparent when the models were applied to an ephemeral, low-yielding catchment. It presented more considerable difficulties for rainfall-runoff model estimation than the humid, temperate catchment where less substantial performance differences were detected.

Second, the analyses presented in this paper consider model performances during model calibrations as well as during simulations. Obviously, model performances in simulation mode are critically important, as the primary motivation for modeling often is prediction. When looking at the results from Salmon Brook catchment, it appears that during simulation the performance difference in favor of the less conceptual model becomes more distinct.

Third, the paper also illustrates why extreme care is necessary if models are used for reproduction of mass fluxes where no measured data are available. The model of a more conceptual nature, which is capable of delivering an estimate for catchment evapotranspiration, is modified by changing the relationship between actual and potential evapotranspiration. When the results yielded by the modified model are compared with those of the original one, it is noted that although estimated streamflows are very much alike, the calculated evapotranspiration time series are drastically different.

Acknowledgments. The first author is grateful to the Finnish Cultural Foundation and The Land and Water Technology Foundation for providing funding that made his stay in the Centre for Resource and Environmental Studies possible. We are grateful for the availability of data from the Coweeta Hydrological Laboratory and the Salmon catchment. The reviewers' comments were very welcome and improved the manuscript.

References

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. L. Rasmussen, An introduction to the European Hydrology System SHE, 2, Structure of a physically-based, distributed modeling system, *J. Hydrol*., *87*, 61–77, 1986.

Bathurst, J. C., Sensitivity analysis of the Système Hydrologique Européen for an upland catchment, *J. Hydrol*., *87*, 103–123, 1986.

Beck, M. B., Forecasting environmental change, *J. Forecast*., *10*, 3–19, 1991.

Beck, M. B., F. M. Kleissen, and H. S. Wheater, Identifying flow paths in models of surface water acidification, *Rev. Geophys*., *28*, 207–230, 1990.

Bergström, S., The HBV model, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, Water Resour. Publ., Highlands Ranch, Colo., 1995.

Beven, K. J., Changing ideas in hydrology: The case of physically-based models, *J. Hydrol*., *105*, 157–172, 1989.

Beven, K. J., A. Calver, and E. M. Morris, The Institute of Hydrology distributed model, *Tech. Rep.* 89, Inst. Hydrol., Wallingford, England, 1987.

Binley, A. M., and K. J. Beven, A physically based model of heterogeneous hillslopes, 2, Effective hydraulic conductivities, *Water Resour. Res*., *25*, 1227–1233, 1989.

Boughton, W. C., A simple model for estimating the water yield of ungauged catchments, in *Civil Engineering Transactions*, pp. 83–88, Inst. of Eng., Canberra, ACT, Australia, 1984.

Chapman, P. J., B. Reynolds, and H. S. Wheater, Hydrochemical changes along stormflow pathways in a small headwater moorland catchment in mid-Wales, U.K., *J. Hydrol*., *151*, 241–265, 1993.

Chapman, T., Optimization of a rainfall-runoff model for an arid zone catchment, in *IAHS-UNESCO Symposium on the Results of Research on Representative and Experimental Basins*, *Publ. 96*, pp. 126–144, IAHS Press, Wallingford, England, 1970.

Chapman, T., and R. Malone, Comparison of models for estimation of groundwater recharge, using data from a deep weighing lysimeter, in *MODSIM 99*, edited by L. Oxley and F. Scrimgeour, pp. 165–170, The Model. and Simul. Soc. of Aust. and New Zealand Inc., Univ. of Waikato, Hamilton, New Zealand, 1999.

Chiew, F., and T. McMahon, Application of the daily rainfall runoff model MODHYDROLOG to 28 Australian catchments, *J. Hydrol*., *153*, 383–416, 1994.

Christophersen, N., and R. F. Wright, Sulphate budget and a model for sulphate concentrations in stream water at Birkenes, a small forested catchment in southernmost Norway, *Water Resour. Res*., *17*, 377–389, 1981.

Cosby, B. J., G. M. Hornberger, J. N. Galloway, and R. F. Wright, Modeling the effects of acid deposition: Assessment of a lumped parameter model of soil water and stream water chemistry, *Water Resour. Res*., *21*, 51–63, 1985.

Crawford, N. H., and R. K. Linsley, Digital simulation in hydrology: Stanford Watershed Model IV, *Dep. of Civ. Eng. Rep. 39*, Stanford Univ., Stanford, Calif., 1966.

Dietrich, C. R., J. P. Norton, and A. J. Jakeman, Ill-conditioning in environmental system modeling, in *Modeling Change in Environmental Systems*, edited by A. J. Jakeman, M. B. Beck, and M. J. McAleer, pp. 37–57, John Wiley, New York, 1993.

Dumenil, L., and E. Todini, A rainfall-runoff scheme for use in the Hamburg climate model, in *Advances in Theoretical Hydrology*, edited by J. P. O'Kane, pp. 129–157, Elsevier Sci., New York, 1992.

Eagleson, P. S., The emergence of global-scale hydrology, *Water Resour. Res*., *22*, 6S–14S, 1986.

Edijatno Nascimento, N. De O., X. L. Yang, Z. Makhlouf, and C. Michel, GR3J: A daily watershed model with three free parameters, *Hydrol. Sci. J.-J. Sci. Hydrol*., *44*, 263–277, 1999.

Evans, J. P., and A. J. Jakeman, Development of a simple, catchment-scale rainfall-evapotranspiration-runoff model, *Environ. Model. Software*, *13*, 385–393, 1998.

Freer, J., K. Beven, and B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data—An application of the GLUE approach, *Water Resour. Res.*, *32*, 2161–2173, 1996.

Freeze, R. A., Role of subsurface flow in generating surface runoff, 2, Upstream source areas, *Water Resour. Res.*, *8*, 1272–1283, 1972.

Gan, T. Y., and S. J. Burges, An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments, 2, Hydrologic responses for normal and extreme rainfall, *Water Resour. Res.*, *26*, 1605–1619, 1990.

Grayson, R. B., I. D. Moore, and T. A. McMahon, Physically based hydrologic modeling, 2, Is the concept realistic?, *Water Resour. Res.*, *28*, 2659–2666, 1992.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, Toward improved calibration of hydrologic models—Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, 1998.

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

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

Johnston, P. R., and D. H. Pilgrim, Parameter optimization for watershed models, *Water Resour. Res.*, *12*, 477–486, 1976.

Kuhl, S. C., and J. R. Miller, Seasonal river runoff calculated from a global atmospheric model, *Water Resour. Res.*, *28*, 2029–2039, 1992.

Leung, L. R., M. S. Wigmosta, S. J. Ghan, D. J. Epstein, and L. W. Vail, Application of a subgrid orographic precipitation surface hydrology scheme to a mountain watershed, *J. Geophys. Res.*, *101*, 12,803–12,817, 1996.

Littlewood, I. G., and A. J. Jakeman, A new method of rainfall-runoff modeling and its applications in catchment hydrology, in *Environmental Modeling*, vol. II, edited by P. Zannetti, pp. 143–171, Comput. Mech., Billerica, Mass., 1994.

Natale, L., and E. Todini, A constrained parameter estimation technique for linear models in hydrology, in *Mathematical Models for Surface Water Hydrology*, edited by T. A. Ciriani, U. Malone, and J. R. Wallis, pp. 109–147, John Wiley, New York, 1977.

Sherman, L. K., Streamflow from rainfall by the unit-hydrograph method, *Eng. News Rec.*, *108*, 501–505, 1932.

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

Spear, R. C., T. M. Grieb, and N. Shang, Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, *30*, 3159–3169, 1994.

Sugawara, M. I., I. Watanabe, E. Ozaki, and Y. Katsuyame, Reference Manual for the TANK Model, report, Nat. Res. Cent. for Disaster Prev., Tokyo, Japan, 1983.

Swift, L. W., Jr., G. B. Cunningham, and J. E. Douglass, Climatology and hydrology, in *Forest Hydrology and Ecology at Coweeta*, edited by W. T. Swank and D. A. Crossley Jr., pp. 33–55, Springer-Verlag, New York, 1988.

Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, A land-surface hydrology parameterization with subgrid variability for general circulation models, *J. Geophys. Res.*, *97*, 2717–2728, 1992.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*, 83–97, 1998.

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

A. J. Jakeman, Center for Resource and Environmental Studies, Australian National University, Canberra, ACT 0200, Australia. (tony@cres.anu.edu.au)

T. S. Kokkonen, Laboratory of Water Resources, Helsinki University of Technology, P.O. Box 5300, Helsinki, FIN-02015 HUT, Finland. (teemu.kokkonen@hut.fi)

(Received May 15, 2000; revised December 7, 2000; accepted May 16, 2001.)