Validation of a watershed model without calibration

Richard M. Vogel

WaterSHED Center, Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts, USA

A. Sankarasubramanian

International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA

Received 23 December 2002; revised 5 June 2003; accepted 22 July 2003; published 18 October 2003.

[1] Traditional approaches for the validation of watershed models focus on the "goodness of fit" between model predictions and observations. It is possible for a watershed model to exhibit a "good" fit, yet not accurately represent hydrologic processes; hence "goodness of fit" can be misleading. Instead, we introduce an approach which evaluates the ability of a model to represent the observed covariance structure of the input (climate) and output (streamflow) without ever calibrating the model. An advantage of this approach is that it is not confounded by model error introduced during the calibration process. We illustrate that once a watershed model is calibrated, the unavoidable model error can cloud our ability to validate (or invalidate) the model. We emphasize that model hypothesis testing (validation) should be performed prior to, and independent of, parameter estimation (calibration), contrary to traditional practice in which watershed models are usually validated after calibrating the model. Our approach is tested using two different watershed models at a number of different watersheds in the United States. *INDEX TERMS:* 1836 Hydrology: Hydrologic budget (1655); 1860 Hydrology: Runoff and streamflow; 1899 Hydrology: General or miscellaneous; 9820 General or Miscellaneous: Techniques applicable in three or more fields; *KEYWORDS:* rainfall-runoff, catchment model, verification, hypothesis testing, moment diagram

Citation: Vogel, R. M., and A. Sankarasubramanian, Validation of a watershed model without calibration, *Water Resour. Res.*, *39*(10), 1292, doi:10.1029/2002WR001940, 2003.

1. Introduction

[2] Since the advent of computer technologies, deterministic watershed model structures have grown increasingly sophisticated and complex. Watershed model structures have evolved from lumped to distributed, requiring vast increases in both input data and model parameter characterization. Numerical solution algorithms have evolved from coarse resolution in time and space to much finer resolution. For example, model time steps have evolved from annual and monthly to daily and hourly, requiring vast increases in computational requirements. Remote sensing technologies and database management systems have improved so that now satellite monitoring systems and geographic information systems are routinely used to monitor and manage distributed data sources. With these developments have come significant improvements in our understanding and ability to calibrate and validate deterministic watershed models. Concurrently, as model structures become more sophisticated and complex, their calibration and validation pose even greater challenges.

[3] The usefulness of watershed models as tools to predict watershed responses remains questionable as does the validity of the internal watershed process dynamics within each model. This is because in most applications, little or no data are available other than the input and output data used to calibrate the model. Consequently, the simulated responses of the watershed (other than streamflows) remain "internal" to the model and are not subject to the type of scientific scrutiny one would expect from such modeling exercises. More recently, attempts have been made to examine the internal processes of some watershed models to aid in model validation [*Seibert et al.*, 1997]. When one considers the wide range of watershed models and the heavy emphasis on their calibration [*Duan et al.*, 2003], it is surprising how little attention has been given to the problem of model validation. In three recent reviews of watershed modeling [*Singh*, 1995; *Hornberger and Boyer*, 1995; *Singh and Woolhiser*, 2002] and watershed model calibration [*Duan et al.*, 2003], there was little attention given to developments in the area of model validation.

[4] Research on watershed modeling has evolved considerably, along with our awareness that model structures and their associated model parameter sets are not unique and that infinite plausible mathematical representations exist. Groundwater modelers [*Konikow and Bredehoeft*, 1992], Earth science modelers [*Oreskes et al.*, 1994], watershed modelers [*Kirchner et al.*, 1996; *Wagener et al.*, 2003], and others [*Oreskes and Belitz*, 2001] have come to the realization that scientists have understood at least since *Popper* [1959], that as hydrologists we can never validate a watershed model hypothesis, only invalidate it! This is similar to a statistician's approach to hypothesis testing. Acceptance or rejection of a statistical model is philosophically identical to acceptance or rejection of a deterministic watershed

0043-1397/03/2002WR001940$09.00

model since all deterministic models contain an unavoidable stochastic component [*Vogel*, 1999]. *Oreskes and Belitz* [2001] argue that the term "validation" is unfortunate because its root, "valid," implies a legitimacy that we are not justified in asserting.

[5] The *U.S. Department of Energy* [1986] defines validation as the determination "that the code or model indeed reflects the behavior of the real world." Similarly, the *International Atomic Energy Agency* [1982] defines a validated model as one which provides "a good representation of the actual processes occurring in a real system." Although there is rough agreement on the goal of model validation, no agreement exists on a uniform methodology for executing model validation.

[6] Traditional approaches to the validation of watershed models are normally variations on the split-sample technique. The traditional idea of validation is to assure that the model can make accurate and reproducible predictions outside the period of time over which it was calibrated [*Klemes*, 1986; *Tsang*, 1991; *Flavelle*, 1992; *Refsgaard and Knudsen*, 1996]. *Klemes* [1986] introduced a hierarchical scheme for the validation of hydrologic models which tests a model's ability to make predictions outside the calibration period (split-sample), on different basins (proxy-basin), and under different climate regimes (differential split-sample). *Refsgaard and Knudsen* [1996] are the only ones we could locate who applied *Klemes*'s [1986] hierarchical approach to model validation.

[7] Traditional approaches to model validation concentrate on the spatial and temporal transposability [*Klemes*, 1986] of models, which are issues central to the application of models. Yet validation exercises could be further generalized to include all approaches which provide grounds for credibility for a given model. The framework of statistical hypothesis testing provides a quantitative framework for determining model credibility or validity [*Hooper*, 2001]. *Luis and McLaughlin* [1992] and *Flavelle* [1992] introduced a statistical framework for the validation of watershed models. Their objectives are similar to those outlined here; however, their approach differs from ours because their hypothesis tests focus on whether or not the model errors are negligible. Therefore their hypothesis tests focus attention on model errors, which implies the model is already fit to data (calibrated), and hence their approach still focuses on the "goodness of fit" of the model to observations, within the context of some prespecified objective function.

[8] Our approach is to augment the current calibration/validation paradigm with a method which does not focus solely upon the traditional "goodness of fit" of the model predictions to the observations. Rather, our approach focuses on the ability of the hypothesized watershed model structure to represent the observed covariance structure of the input and output time series without ever calibrating the model. Since our approach focuses on the modeled covariance structure of the input and output series, we term our approach a covariance approach to model validation. In the following section we demonstrate our covariance approach using a simple linear watershed model for which the covariances can be derived analytically. Following that example, we demonstrate the application of our covariance approach to a more realistic nonlinear watershed model which requires a numerical approach to estimate the covariances.

2. An Analytical Example of the Covariance Approach to Watershed Model Validation

[9] In this section we introduce a simple linear watershed model for the purpose of demonstrating how covariances can be employed for the validation of a watershed model. Since the model is linear, the covariances are derived analytically. After introducing the model and deriving various covariances, we introduce our proposed validation methodology and demonstrate some of its advantages over existing validation methods.

2.1. The "abc" Model

[10] The "abc" model, originally conceived by Harold A. Thomas, was introduced by *Fiering* [1967] as a pedagogic tool for modeling relationships among precipitation, evapotranspiration, groundwater storage, and streamflow using only three model parameters. Since the model is linear and lacks a soil moisture component, it is not expected to perform well, yet consistent with *Fiering*'s [1967] original intent, this model provides a wealth of pedagogic opportunities for demonstrating fundamental hydrologic and statistical issues associated with watershed modeling. Other applications of the abc model include studies by *Salas and Smith* [1981], *Kuczera* [1982], and *Rogers and Fiering* [1990]. The linear structure of the abc model enables us to derive analytical relationships between moments of the input (precipitation), output (streamflow), model error, and model parameters. When such analytical relations are unavailable, one may resort to the use of numerical methods, as is discussed in the next section. The analytical moments of the abc model derived below differ from previous studies, because previous derivations of the modeled moments [*Fiering* 1967; *Rogers and Fiering*, 1990] ignored model error.

[11] The abc model is a water balance defined by a continuity equation for the surface and saturated groundwater components of the hydrologic cycle. If precipitation at time t is represented by P_t, then infiltration is $I_t = aP_t$ and evapotranspiration is $E_t = bP_t$, where a and b represent the fraction of rainfall which infiltrates and evaporates, respectively. The remaining component of rainfall, $P_t - I_t - E_t = (1 - a - b)P_t$, results in surface runoff to the stream channel. Groundwater storage at time t is G_t, and the groundwater outflow to the stream channel is a fixed fraction cG_{t-1} of groundwater storage in the previous period. Finally, streamflow, Q_t, is given as the combination of surface and groundwater inputs in addition to model error ε_t

$$Q_t = (1 - a - b)P_t + cG_{t-1} + \varepsilon_t \tag{1}$$

Groundwater storage G_t is derived by continuity as previous groundwater storage G_{t-1} less groundwater outflow plus infiltration and model error ν_t

$$G_t = (1 - c)G_{t-1} + aP_t + \nu_t \tag{2}$$

The model is named after its three parameters, a,b, and c, that are presumed to have some degree of physical interpretation. Since the parameters represent fractions, they have upper and lower limits $0 \leq a, b, c \leq 1$, and since infiltration and evapotranspiration combined cannot exceed total precipitation, $0 \leq a + b \leq 1$.

2.2. Moments of the "abc" Model

[12] Ignoring the error terms ε_t and ν_t, *Fiering* [1967] and *Rogers and Fiering* [1990] derived the mean and variance of streamflow from equations (1) and (2) for the steady state case where $E[G_t] = E[G_{t-1}]$. Model error plays a fundamental role in model validation and cannot be ignored. *Kuczera* [1982] shows how important it is to account for, rather than ignore, the errors in equations (1) and (2), particularly when one's interest is in model parameter estimation. Dropping the subscripts, one can show that the mean, variance, and lag-one serial correlation of streamflow are given by

$$E[Q] = \mu_Q = (1-b)\mu_P \tag{3}$$

$$\mathrm{Var}[Q] = \sigma_Q^2 = \sigma_P^2\left[(1-a-b)^2 + \frac{ca^2}{2-c}\right] + \left[\frac{c}{2-c}\right]\sigma_\nu^2 + \sigma_\varepsilon^2 \tag{4}$$

$$\rho_1 = \frac{\left[ac(1-a-b) + \left[\frac{a^2c(1-c)}{2-c}\right]\right]\sigma_P^2 + \left[\frac{(1-c)c}{2-c}\right]\sigma_\nu^2}{\sigma_Q^2} \tag{5}$$

where μ_p and σ_p^2 are the mean and variance of the precipitation, respectively, and σ_ε^2 and σ_ν^2 are the variance of the model residuals. In this initial example, the residuals are assumed to have zero mean and to be independent of one another. When $\sigma_\varepsilon^2 = 0$ and $\sigma_\nu^2 = 0$, equations (4) and (5) reduce to the expressions derived by *Fiering* [1967] and *Rogers and Fiering* [1990], which ignored those error terms. One can also show that the correlation between the input (precipitation) and output (streamflow) is given by

$$\rho(P,Q) = \frac{(1-a-b)\sigma_P}{\sigma_Q} \tag{6}$$

[13] We assume groundwater measurements are unavailable, and hence it is impossible to estimate σ_ν^2. Instead, we make an assumption about the relationship between the residuals ε_t and ν_t. We assume the relative prediction error associated with streamflow Q_t is the same as the relative prediction error associated with groundwater storage G_t, i.e., $\sigma_\varepsilon/\mu_Q = \sigma_\nu/\mu_G$. Since $\mu_Q = (1-b)\mu_P$ and $\mu_G = a\mu_P/c$, we obtain the relationship

$$\frac{\sigma_\varepsilon}{\sigma_\nu} = \frac{(1-b)c}{a} \tag{7}$$

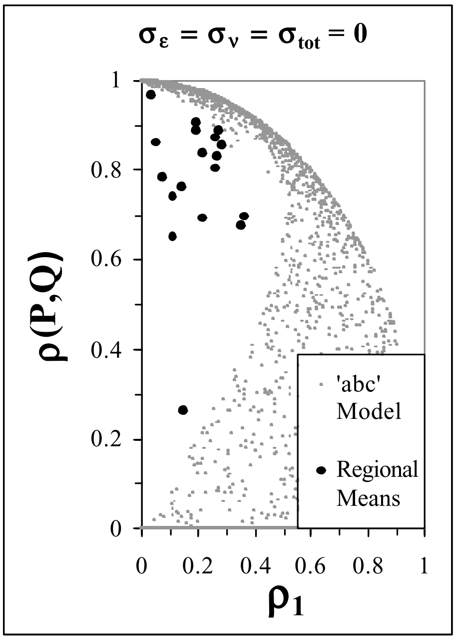
Finally, we assume that the ground and surface-water errors are uncorrelated so that the total model error variance is the sum of the variance of the ground and surface water model errors

$$\sigma_{\mathrm{tot}}^2 = \sigma_\varepsilon^2 + \sigma_\nu^2 \tag{8}$$

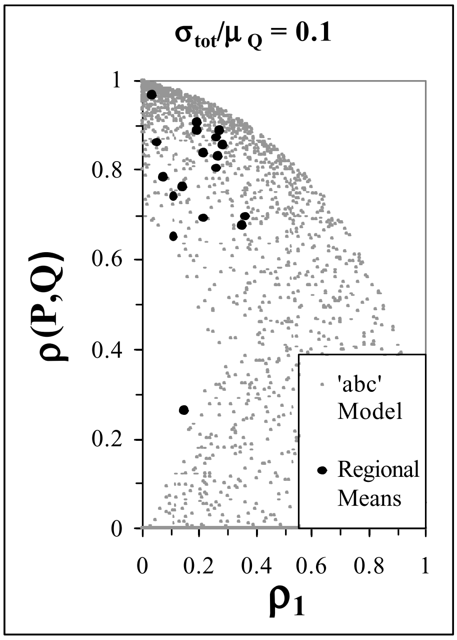
2.3. Use of Covariances to Invalidate the "abc" Watershed Model

[14] Techniques for the validation of watershed models should reflect the goals of the modeling exercise. If the goal is to reproduce peak discharges, the model performance criteria should reflect that goal. We begin by using covariances to evaluate a watershed model with respect to a desired modeling goal, although other statistics may be used such as low flow or flood peak discharges. Assume the goal is for the model to reproduce the observed lag-one serial correlation of the output (streamflow), ρ_1, and the observed cross correlation between the input (precipitation) and output (streamflow), $\rho(P, Q)$, using a yearly time step. This goal might be important for long-term water supply planning, where the correlation structure of the model output is important to maintain. This idea is illustrated in Figure 1. Figure 1 is constructed in such a way as to illustrate the complete sample space of values of ρ_1 and $\rho(P, Q)$ that the model is capable of reproducing using an annual time step. Shown in shaded dots are the results of 5000 Monte Carlo experiments which used equations (3)–(8) to compute the relationship between ρ_1 and $\rho(P, Q)$. Those shaded dots are obtained by sampling the model parameters a, b, and c from a uniform distribution over the interval [0,1] (with $0 < a + b \leq 1$) so as to represent all possible modeled moments corresponding to all possible combinations of model parameters. This parameter generation scheme was done for simplicity; in practice, one might constrain model parameters to ensure realistic parameter values. Similarly, we sample the coefficient of variation of streamflow and precipitation from a uniform distribution over the range of values of these statistics observed for annual time series across the entire United States [*Vogel et al.*, 1998]. The net result is the region formed by the shaded dots in Figure 1a, which illustrates the possible sample space of these two covariances which this particular model can reproduce, without model error ($\sigma_\varepsilon = \sigma_\nu = \sigma_{\mathrm{tot}} = 0$). For comparison, we plot regional average sample values of these two statistics using large solid circles. The regional average values of ρ_1 were obtained from *Vogel et al.* [1998], and a similar approach was used to obtain regional values of $\rho(P, Q)$ for the 18 major water resource regions of the United States. The regional sample statistics illustrated in Figure 1 summarize the properties of these two covariances at 1557 watersheds across the United States based on U.S. Geological Survey streamflow records [*Slack et al.*, 1993]. It is important to stress that the regional sample estimates of moments illustrated in Figure 1 do not depend upon any model assumptions, and since they are based on regional information, they also have very small sampling error. Since the abc model is unable to reproduce the observed moments, we can reject the abc model at the annual timescale for all regions of the United States.

[15] Figure 1b is constructed differently from Figure 1a because it assumes a slight model error exists ($\sigma_{\mathrm{tot}}/\mu_Q = 0.1$). Model error is always introduced when a model is calibrated to data. In Figure 1b, it now appears that the abc model can reproduce the necessary moments. This experiment documents how the introduction of model error (which normally occurs during the calibration process) can confuse us into thinking a model is acceptable, when in fact it is not. This experiment also documents how promising this approach can be, because in this instance, a single graphic image was created to invalidate the use of an annual abc watershed model, for the chosen goal, for all regions of the United States.



(a)



(b)

Figure 1. Shaded dots depict relationship between cross correlation of annual precipitation and streamflow, $\rho(P, Q)$, and lag-one serial correlation of annual streamflow ρ_1 corresponding to nearly all possible combinations of abc model parameters. Solid circles denote the regional estimates of the moments $\rho(P, Q)$ and ρ_1 based on 1337 watersheds across the continental United States.

[16] The Monte Carlo approach taken in Figure 1 to develop the complete sample space of moments that the model is capable of representing is analogous to the generalized sensitivity analysis introduced by *Spear and Hornberger* [1980] and used subsequently by others for the purpose of evaluating the sensitivity of a goodness-of-fit criterion to all possible combinations of model parameters. The methodology outlined above is unique because the abc model is linear, enabling us to derive analytical expressions for various moments of streamflow as a function of moments of precipitation. Further, the use of an annual timescale enabled us to generalize the behavior of streamflow and precipitation for all regions of the United States. In the following section we illustrate our approach in a more realistic setting in which the watershed model is nonlinear and the timescale is monthly.

3. An Example of the Covariance Approach to Validation of a Monthly Watershed Model

[17] In this section we introduce a more realistic nonlinear watershed model than in the previous example and document how our covariance approach to model validation would be applied in a more realistic situation at a single site.

3.1. The "abcd" Model

[18] In contrast to the linear abc model, which only accepts precipitation as input and only models groundwater, the abcd model is a nonlinear water balance model which accepts precipitation and potential evapotranspiration as inputs and captures the mechanics of soil moisture, saturated groundwater, and streamflow. The abcd model was originally introduced by *Thomas* [1981] and applied using an annual time step. *Sankarasubramanian and Vogel* [2002] evaluated the goodness-of-fit of an annual abcd model to 1337 watersheds across the United States. *Alley* [1984] and *Vandeweile et al.* [1992] found that a monthly abcd model compared favorably with several other monthly water balance models. *Fernandez et al.* [2000] evaluated the performance of the monthly abcd model on 33 watersheds in the southeastern United States and introduced a regional approach to estimation of the model parameters. Since *Thomas* [1981], *Alley* [1984], *Fernandez et al.* [2000], and *Sankarasubramanian and Vogel* [2002] summarize the abcd model, we do not reproduce the model structure here.

3.2. Use of Covariances to Invalidate the Annual "abcd" Model

[19] We begin by applying our validation approach to an annual version of the abcd model, and then in the following section we explore a monthly version of the same model. Since the abcd model is a nonlinear model, it is not possible to derive exact analytical expressions for various moments of model output, as was the case above for the abc model. However, it is always possible to estimate covariances between model input and output for different model parameter combinations using computer simulation. In this section we apply the covariance approach described in the previous section to the abcd model at the Coosawhatchi River near Hampton, South Carolina (USGS Site 02176500) and the St. Johns River near Deland, Florida (USGS Site 02236000). Streamflow data Q for these two sites are obtained from the U.S. Geological Survey [*Slack et al.*, 1993] over the 37-year period 1951–1988. Appendix A summarizes the procedures employed to estimate the monthly and annual input time series of precipitation P and potential evapotranspiration PE required for the following experiments.

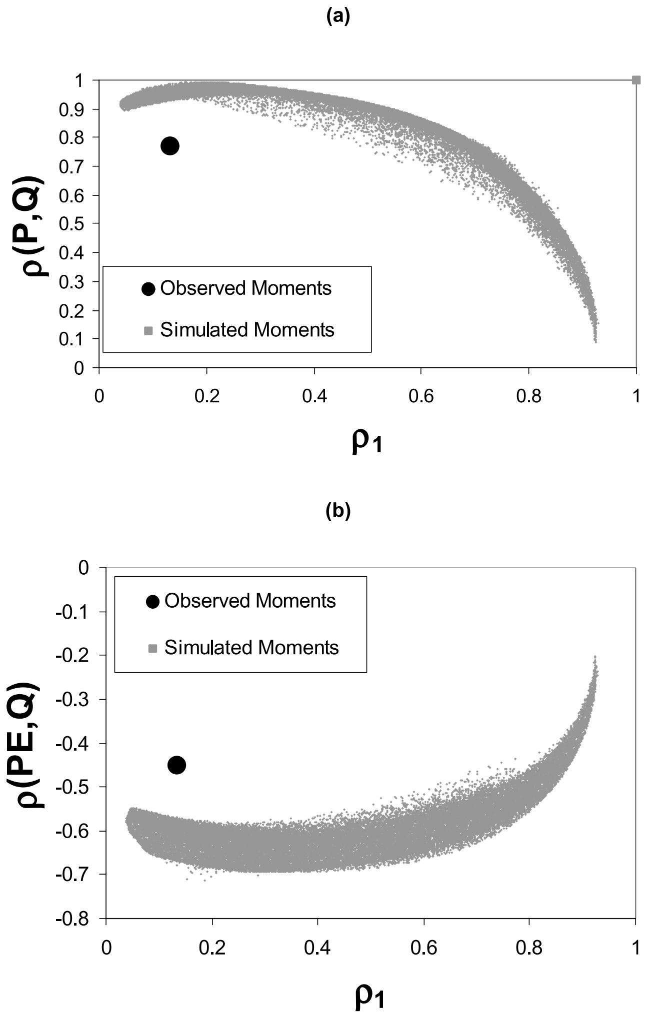


Figure 2. Comparison of simulated and observed relationships between (a) cross correlation of annual precipitation and streamflow, $\rho(P, Q)$ and lag-one serial correlation of streamflow ρ_1 and (b) cross correlation of annual potential evapotranspiration and streamflow, $\rho(PE, Q)$ and lag-one serial correlation of annual streamflow ρ_1 for the Coosawhatchie River watershed near Hampton, South Carolina, using an annual abcd model.

[20] Figures 2 and 3 compare the simulated and observed relationships between the cross correlation of annual precipitation and streamflow, $\rho(P, Q)$, and lag-one serial correlation of streamflow, ρ_1, and between the cross correlation of annual potential evapotranspiration and streamflow, $\rho(PE, Q)$, and ρ_1 for the Coosawhatchie River and St. Johns River watersheds, respectively. The simulated moments in Figures 2 and 3 are based on 50,000 independent sets of parameters of the abcd model. The model parameters *a*, *c*, and *d* are generated from a uniform (0, 1) distribution, and the model parameter *b* is generated over the range (0, 2000) mm. The parameter *b* is the upper limit on the sum of soil moisture and evapotranspiration, and hence an upper bound for *b* would be about twice the value of mean annual precipitation. An upper bound of $b = 2000$ was used for both watersheds. For each of the 50,000 generated abcd model parameter sets, 50,000 corresponding sets of streamflow records *Q* were generated using the observed annual time series of *P* and *PE* for each watershed. For each of the 50,000 generated sets of *P*, *PE*, and *Q*, 50,000 corresponding sets of simulated covariances $\rho(P, Q)$, $\rho(PE, Q)$, and ρ_1 were computed and reported in Figures 2 and 3 using shaded dots. Also shown in Figures 2 and 3 are the observed sample covariances $\rho(P, Q)$, $\rho(PE, Q)$, and ρ_1 shown using large solid circles. Figures 2 and 3 document that the abcd model is unable to reproduce relationships between the observed covariances. In Figures 2a, 2b, and 3b, the observed moments $\rho(P, Q)$, $\rho(PE, Q)$, and ρ_1 lie outside the sample space of covariances which the annual abcd model is capable of reproducing. This experiment documents that an annual abcd model cannot reproduce the observed annual covariance structure between the input and output time series at either watershed. Hence we conclude that an annual abcd model is unable to capture the basic covariance structure of annual climate and stream-

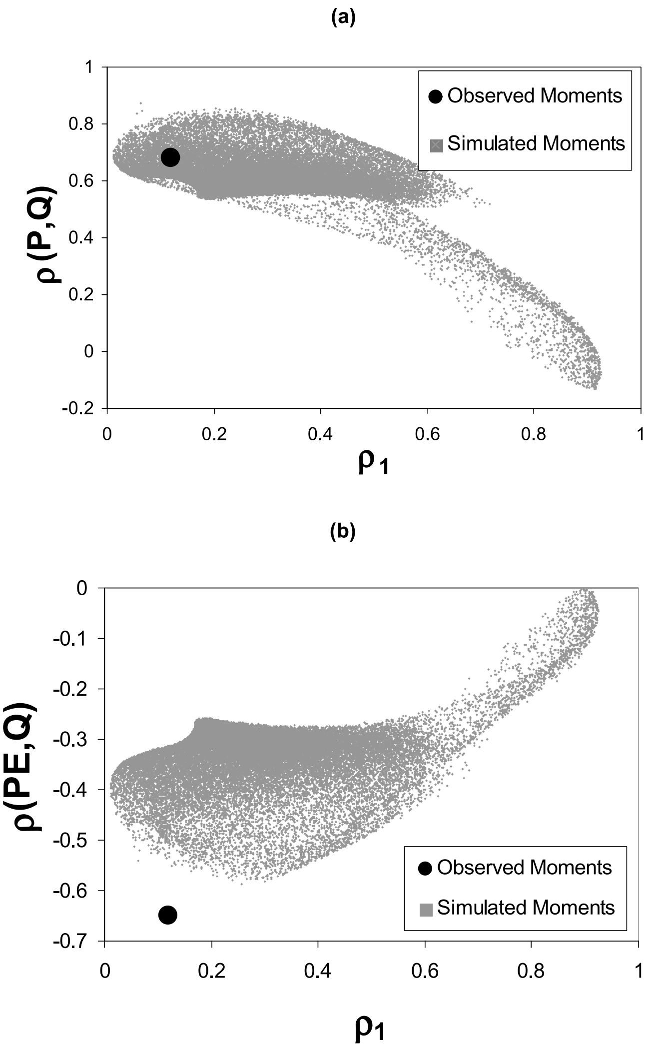


Figure 3. Comparison of simulated and observed relationships between (a) cross correlation of annual precipitation and streamflow, $\rho(P, Q)$, and lag-one serial correlation of streamflow ρ_1 and (b) cross correlation of annual potential evapotranspiration and streamflow, $\rho(PE, Q)$, and lag-one serial correlation of annual streamflow ρ_1 for the St. Johns River watershed near Deland, Florida, using an annual abcd model.

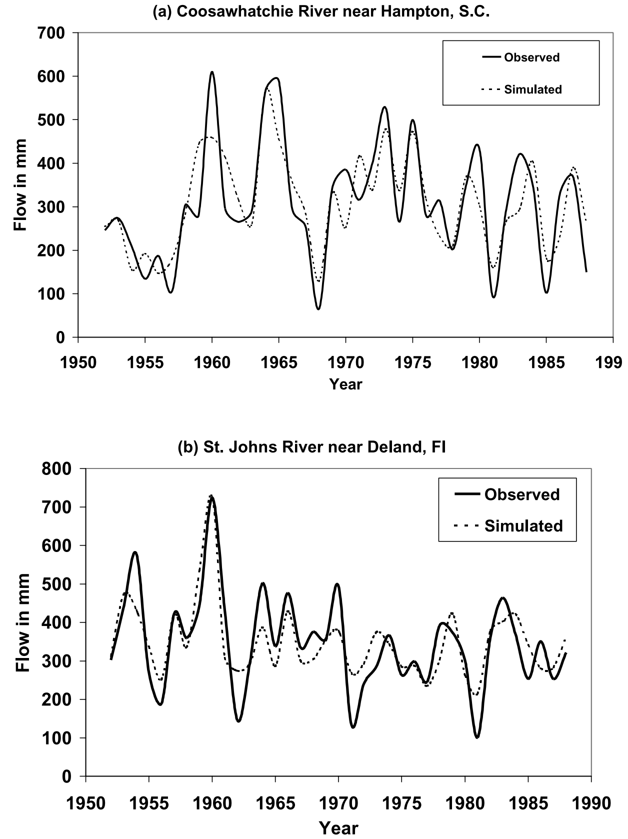


Figure 4. Comparison of the observed annual flow series with the modeled annual flow series based on a fitted annual abcd model at both watersheds.

flow series at these two watersheds. We conclude that we can reject the abcd model at the annual timescale for these two watersheds.

[21] Note, however, that when one views the goodness of fit of an annual abcd model, determination of the adequacy or inadequacy of the abcd model is not nearly as definitive as when one uses the covariance approach to model validation. The traditional "goodness of fit" approach is illustrated in Figure 4, which compares the observed annual flow series with the modeled annual flow series at both watersheds based on a calibrated annual abcd model. Calibration of the abcd model is performed using the shuffled complex evolution (SCE) algorithm developed by *Duan et al.* [1992] for the calibration of watershed models. The bias in reproducing the mean annual flows at both stations is negligible, and the correlation between the observed and simulated annual flows at both the stations is 0.81. On the basis of a traditional goodness-of-fit criterion along with Figure 4, it would be difficult to reject this model. The validation approach introduced here provides a much more objective criterion for acceptance or rejection of a model hypothesis than standard goodness-of-fit evaluations of the type illustrated in Figure 4.

3.3. Use of Covariances to Validate the Monthly "abcd" Model

[22] One should not be surprised that by using an annual time step, we were able to invalidate both the linear abc and the nonlinear abcd models. This is because fundamental hydrologic processes which describe annual hydrology occur at seasonal, monthly, or shorter timescales. In this section we explore the ability of a monthly abcd model to capture observed relationships among key covariances. The same procedure employed in the previous section was repeated using a monthly abcd model. To summarize the results using the same type of diagram employed in Figures 2 and 3, the 50,000 generated monthly flow traces were aggregated to the annual level to produce 50,000 sets of aggregated annual streamflow traces. The results are illustrated in Figures 5 and 6. Figures 5 and 6 evaluate the ability of a monthly abcd model to reproduce the observed annual covariances $\rho(P, Q)$, $\rho(PE, Q)$, and ρ_1. In Figures 5 and 6 the observed moments $\rho(P, Q)$, $\rho(PE, Q)$, and ρ_1 lie either near the boundary or well within the sample space of covariances which the monthly abcd model is capable of reproducing. This experiment documents that a monthly abcd model is capable of reproducing the observed annual covariance structure between the input and output time series at either of these two watersheds. We conclude that we cannot reject

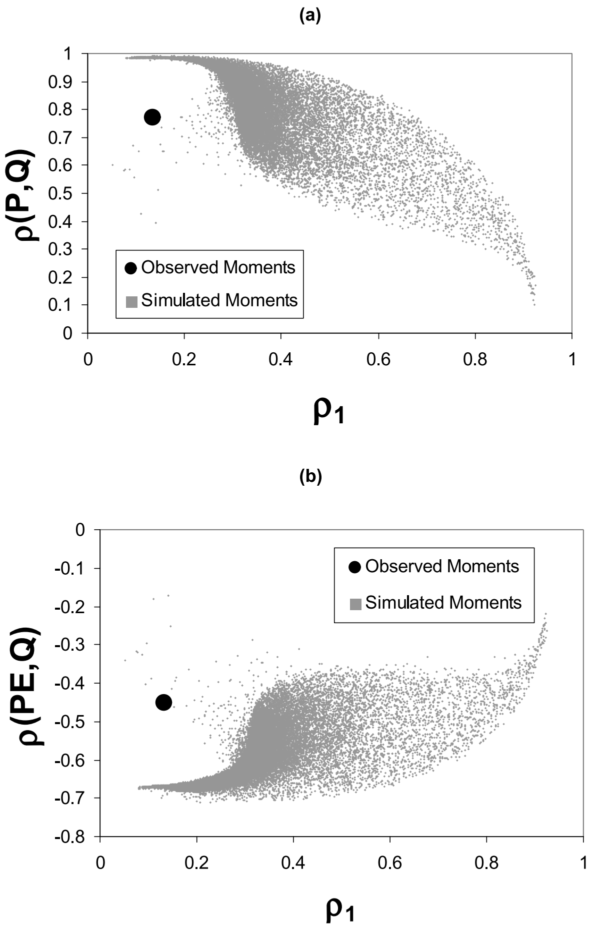


Figure 5. Comparison of simulated and observed relationships between (a) cross correlation of annual precipitation and streamflow, $\rho(P, Q)$ and lag-one serial correlation of streamflow ρ_1 and (b) cross correlation of annual potential evapotranspiration and streamflow, $\rho(PE, Q)$ and lag-one serial correlation of annual streamflow ρ_1 for the Coosawhatchie River watershed near Hampton, South Carolina, using a monthly abcd model.

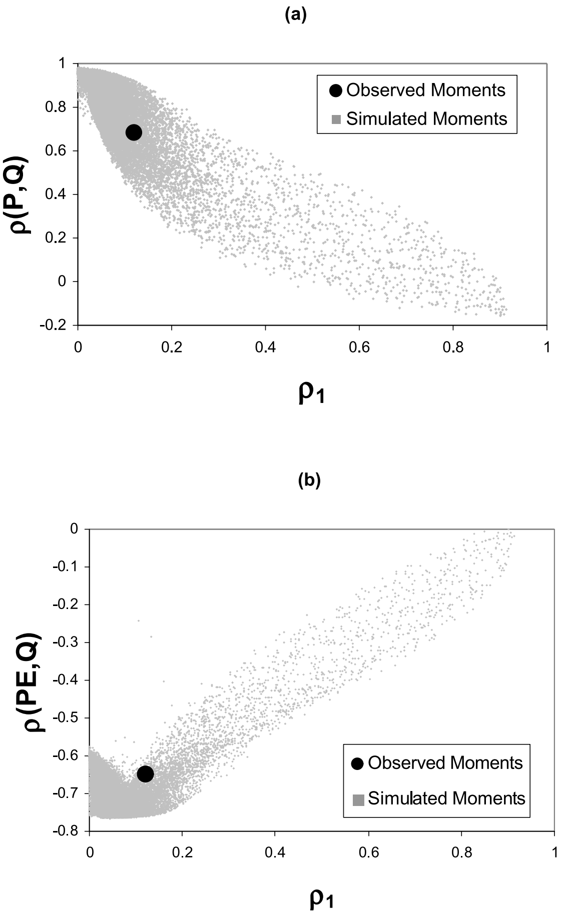


Figure 6. Comparison of simulated and observed relationships between (a) cross correlation of annual precipitation and streamflow, $\rho(P, Q)$, and lag-one serial correlation of streamflow ρ_1 and (b) cross correlation of annual potential evapotranspiration and streamflow, $\rho(PE, Q)$, and lag-one serial correlation of annual streamflow ρ_1 for the St. Johns River watershed near Deland, Florida, using a monthly abcd model.

the monthly abcd model for modeling annual flows at either of these two watersheds. In Figure 5 the sample space of the modeled moments is relatively sparse near the boundaries, leading to tentative conclusions for that site. Further sampling experiments could be performed to clarify the boundaries of the model moments.

[23] The above analyses ignore the sampling variability associated with the estimated covariances based on the 37-year hydrologic sequences. A further extension to our approach would be the addition of a confidence ellipse surrounding the solid circles in Figures 2, 3, 5, and 6, which would account for the sampling variability associated with the small sample estimates of the reported covariances.

[24] Again it is instructive to compare our validation approach with the traditional goodness-of-fit approach. Figure 7 compares the observed annual flow series with the modeled annual flow series at both watersheds based on a calibrated monthly abcd model. Calibration of the monthly abcd model is again performed using the SCE algorithm. Now the goodness of fit of the calibrated monthly abcd model is excellent and a significant improvement over the results displayed earlier in Figure 4 for the annual abcd model. These are exactly the type of results one expects when a model performs as expected; that is, it is not easily invalidated and exhibits an excellent goodness of fit.

4. Conclusions

[25] Any hydrologist could make plausible, yet tentative and not entirely definitive arguments why a particular watershed model is or is not an adequate representation of reality. Yet still no definitive quantitative method or criteria exist for rejecting a watershed model. The validation approach outlined here offers a quantitative methodology for accepting or rejecting a watershed model. Our methodology involves use of generalized sensitivity analysis to explore the ability of a watershed model to reproduce key statistical characteristics associated with the input and output data which are to be used later on to calibrate the model. Although our approach focuses initially on the ability of the model to reproduce key covariances among input and output traces, our approach can easily be adapted to consider the ability of a model to reproduce statistics which are not covariances such as peak flood flows, low flows, flow volumes, or other statistics important for a particular model application.

[26] Our methodology was initially applied to a linear annual abc watershed model at 1337 watersheds across the continental United States. Those experiments revealed that

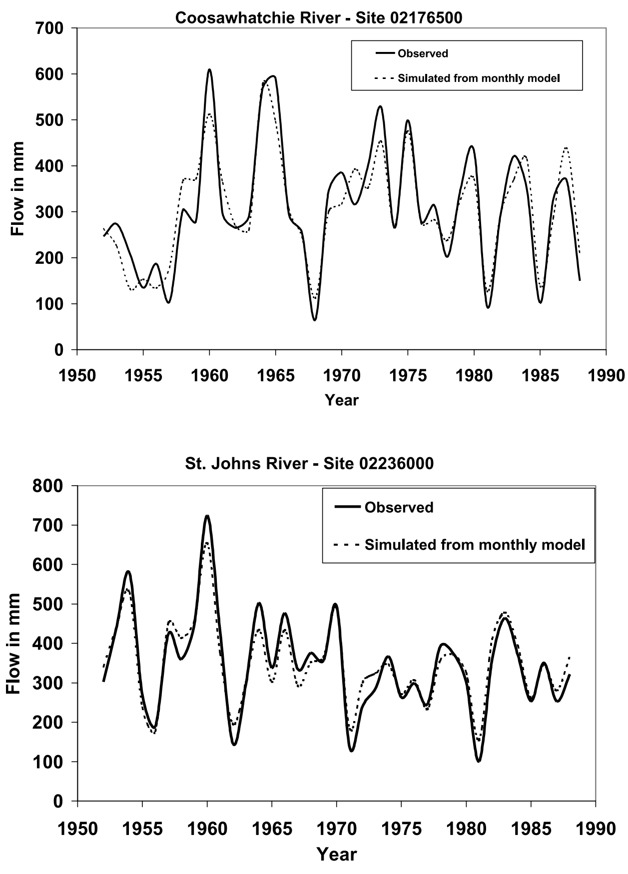


Figure 7. Comparison of the observed annual flow series with the modeled annual flow series based on a fitted monthly abcd model at both watersheds.

an annual abc model is unable to reproduce key covariances between precipitation and streamflow as well as the lag-one correlation of the streamflow; hence we were able to reject the abc model at an annual time step, for the entire United States. This is a very powerful conclusion which would have been difficult to reach using traditional goodness-of-fit evaluations.

[27] In the next experiment we used our methodology to test a nonlinear abcd model, using both monthly and annual time steps, at two watersheds in the southeastern United States. Similar to the results for the annual abc model, it was found that the annual abcd model could not reproduce observed covariances between precipitation *P*, potential evapotranspiration *PE*, and streamflow *Q*, and hence we were able to reject the annual abcd model at both of these sites. By comparison, a traditional graphical goodness-of-fit evaluation (Figure 4) which compared observed and calibrated values of *Q* did not lead to such a definitive conclusion. Finally, it was found that a monthly abcd model could generate monthly streamflows which, when aggregated to the annual level, could reproduce observed covariances between *P*, *PE*, and *Q*. Hence we were unable to reject (or invalidate) a monthly abcd model for modeling annual flows at these two sites.

[28] The validation method outlined here is analogous to the approach used by stochastic streamflow modelers when they perform evaluations which test the ability of a stochastic model to reproduce key statistical characteristics of the data which they are intended to mimic [*Stedinger and Taylor*, 1982]. Our proposed validation method can be implemented both analytically and numerically for any class of hydrologic models, deterministic or stochastic. Both deterministic and stochastic models contain unavoidable model error terms in addition to deterministic terms, and hence they are not as different as one might think [*Vogel*, 1999].

[29] An important goal of this study is to educate hydrologists to view the calibration and validation of deterministic watershed models similarly to the classical statistical problems of parameter estimation and hypothesis testing, respectively. A statistician would not attempt to estimate model parameters (calibration) prior to model hypothesis testing (validation), yet hydrologists routinely calibrate their models prior to validation. Model hypothesis testing (validation) should be performed prior to, and independent of, parameter estimation (calibration), yet deterministic models are usually validated after calibrating the model and observing the goodness of fit of the estimated model. *Hosking and Wallis* [1997] and others have described the steps in the development of statistical hydrologic models documenting clearly that the acceptance or rejection of a hydrologic model (model validation) should be based on a goodness-of-fit test or a hypothesis test which does not depend upon parameter estimation, and only after that step is complete should one begin model parameter estimation (calibration) as the final step in the model building exercise. *Young* [2001] has described the same separation of modeling steps into (1) model identification followed by (2) model parameter estimation for both statistical and physically based hydrologic models. *Hooper* [2001] argues that traditional goodness of fit between observations and predictions is a necessary, but not sufficient, condition for accepting a watershed model hypothesis.

[30] Our hope is that future studies will extend our covariance approach by (1) considering other important statistics which are not necessarily covariances and (2) considering the sampling properties of the sample covariance estimators so that a formal statistical hypothesis test can be implemented instead of the approximate graphical approach used in this initial study.

[31] The covariance validation procedures outlined here can be applied to more complex watershed models than considered here, although not without additional challenges. In this initial study we have evaluated watershed models at the annual timescale, to minimize the number of graphics. In practice, one may wish to evaluate a model's ability to reproduce the observed covariance structure at subannual timescales. It is also unclear a priori how many, or which, covariance statistics to include in the analysis. As more statistics are included, it is more likely that the model will be invalidated. Similarly, as the number of model parameters increases, the more difficult it will be to invalidate such models. Increasingly, the trend is toward use of distributed watershed models with distributed inputs, parameters, and outputs. Selection of appropriate covariance statistics for validation of such distributed models will be even more challenging. Another challenge will be the development of suitable sampling strategies for the model parameter space associated with distributed watershed models.

Appendix A: Climate Database Used for "abcd" Model Experiments

[32] This section describes the procedures used to develop the monthly and annual time series of precipitation *P* and potential evapotranspiration *PE*, required as input to the abcd model. Monthly time series of precipitation, average minimum, and average maximum daily temperature over the period 1951–1988 were obtained from 0.5° time series grids based on the precipitation-elevation regressions on independent slope model (PRISM) climate analysis system [*Daly et al.*, 1994]. PRISM uses a precipitation-elevation regression relationship to distribute point measurements to evenly spaced grid cells. These time series grids were spatially averaged over both basins using a geographic information system. To accomplish this task, the watershed boundaries were delineated using the GTOPO30 global digital elevation model.

[33] Using the monthly time series of average minimum and average maximum temperature data along with extraterrestrial solar radiation for each basin, estimates of monthly potential evapotranspiration were obtained using a method introduced by *Hargreaves and Samani* [1982]. Extraterrestrial solar radiation was estimated for each basin by computing the solar radiation over 0.1° grids using the method introduced by *Duffie and Beckman* [1980] and then summing those estimates over the entire basin.

[34] **Acknowledgments.** The authors are grateful for the encouragement and insights provided by Richard P. Hooper, Charles N. Kroll, George Kuzcera, and John Schaake based on early drafts of this manuscript. The authors are also indebted to three anonymous reviewers for their helpful comments. The initial experiments based on the abc model were presented as a poster titled "The validation of a watershed model" at the 28 July 1999 IUGG workshop on the regionalization of parameters of hydrological and atmospheric land surface models in Birmingham, England. The authors are indebted to Chris Daly for providing them with the monthly time series

grids of temperature and precipitation and to Ian Wilson for his assistance in processing those grids.

References

Alley, W. M., On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models, *Water Resour. Res.*, *20*(8), 1137–1149, 1984.

Daly, C., R. P. Neilson, and D. L. Philips, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*(2), 140–158, 1994.

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

Duan, Q., H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte (Eds.), *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, AGU, Washington, D. C., 2003.

Duffie, J. A., and W. A. Beckman, *Solar Engineering of Thermal Processes*, pp. 1–109, John Wiley, Hoboken, N. J., 1980.

Fernandez, W., R. M. Vogel, and A. Sankarasubramanian, Regional calibration of a watershed model, *Hydrol. Sci. J.*, *45*(5), 689–707, 2000.

Fiering, M. B., *Streamflow Synthesis*, Harvard Univ. Press, Cambridge, Mass., 1967.

Flavelle, P., A quantitative measure of model validation and its potential use for regulatory purposes, *Adv. Water Resour.*, *15*, 5–13, 1992.

Hargreaves, G. H., and Z. A. Samani, Estimating potential evapotranspiration, *J. Irrig. Drain. Eng.*, *108*(3), 225–230, 1982.

Hooper, R. P., Applying the scientific method to small catchment studies: A review of the Panola Mountain experience, *Hydrol. Processes*, *15*, 2039–2050, 2001.

Hornberger, G. M., and E. W. Boyer, Recent advances in watershed modeling, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys.*, *33*, 949–957, 1995.

Hosking, J. R. M., and J. R. Wallis, *Regional Frequency Analysis*, pp. 9–11, Cambridge Univ. Press, New York, 1997.

International Atomic Energy Agency, Radioactive waste management glossary, *Rep. IEA-TEDCDOC-264*, Vienna, 1982.

Kirchner, J. W., R. P. Hooper, C. Kendall, C. Neal, and G. Leavesley, Testing and validating environmental models, *Sci. Total Environ.*, *183*, 33–47, 1996.

Klemes, V., Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*, 13–24, 1986.

Konikow, L. F., and J. D. Bredehoeft, Ground-water models cannot be validated, *Adv. Water Resour.*, *15*, 75–83, 1992.

Kuczera, G., On the relationship between the reliability of parameter estimates and hydrologic time series data used in calibration, *Water Resour. Res.*, *18*(1), 146–154, 1982.

Luis, S. J., and D. McLaughlin, A stochastic approach to model validation, *Adv. Water Resour.*, *15*(1), 15–32, 1992.

Oreskes, N., and K. Belitz, Philosophical issues in model assessment, in *Model Validation: Perspectives in Hydrological Science*, edited by M. G. Anderson and P. D. Bates, chap. 3, pp. 23–41, John Wiley, Hoboken, N. J., 2001.

Oreskes, N., K. Shader-Frechette, and K. Belitz, Verification, validation and confirmation of numerical models in the earth sciences, *Science*, *263*, 641–646, 1994.

Popper, K., *The Logic of Scientific Discovery*, HarperCollins, New York, 1959.

Refsgaard, J. C., and J. Knudsen, Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*(7), 2189–2202, 1996.

Rogers, P. R., and M. B. Fiering, From flow to storage, in *Climate Change and U. S. Water Resources*, edited by P. E. Waggoner, chap. 9, pp. 207–222, John Wiley, Hoboken, N. J., 1990.

Salas, J. D., and R. A. Smith, Physical basis of stochastic models of annual flows, *Water Resour. Res.*, *17*(2), 430–482, 1981.

Sankarasubramanian, A., and R. M. Vogel, Annual hydroclimatology of the United States, *Water Resour. Res.*, *38*(6), 1083, doi:10.1029/2001WR000619, 2002.

Seibert, J., K. H. Bishop, and L. A. Nyberg, A test of TOPMODEL's ability to predict spatially distributed groundwater levels, *Hydrol. Processes*, *11*, 1131–1144, 1997.

Singh, V. P., *Computer Models of Watershed Hydrology*, 1129 pp., Water Resour. Publ., Highlands Ranch, Colo., 1995.

Singh, V. P., and D. A. Woolhiser, Mathematical modeling of watershed hydrology, *J. Hydrol. Eng.*, *7*(4), 270–292, 2002.

Slack, J. R., A. M. Lumb, and J. M. Landwehr, Hydroclimatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variation, 1874–1988, *U.S. Geol. Surv. Water Resour. Invest. Rep., 93-4076*, 193 pp., 1993.

Spear, R. C., and G. M. Hornberger, Eutrphication in Peel Inlet: II. Identification of critical uncertainities via generalized sensitivity analysis, *Water Res.*, *14*, 43–49, 1980.

Stedinger, J. R., and M. R. Taylor, Synthetic streamflow generation: 1. Model verification and validation, *Water Resour. Res.*, *18*(4), 909–918, 1982.

Thomas, H. A., Improved methods for national water assessment, report, contract WR 15249270, U.S. Water Resour. Counc., Washington, D. C., 1981.

Tsang, C.-F., The modelling process and model validation, *Groundwater*, *29*(6), 825–831, 1991.

U.S. Department of Energy, Environmental assessment: Yucca Mountain site, Nevada research and development area, Nevada, vol. 2, *IAEA-TEDCDOC-264*, Off. of Civ. Radioactive Waste Manage., Washington, D. C., 1986.

Vandeweile, G. L., C.-Y. Xu, and N.-L. Win, Methodology and comparative study of monthly water balance models in Belgium, China, and Burma, *J. Hydrol.*, *134*, 315–347, 1992.

Vogel, R. M., Stochastic and deterministic world views, editorial, *J. Water Resour. Plann. Manage.*, *125*(6), 311–313, 1999.

Vogel, R. M., Y. Tsai, and J. F. Limbrunner, The regional persistence and variability of annual streamflow in the United States, *Water Resour. Res.*, *34*(12), 3445–3459, 1998.

Wagener, T., H. S. Wheater, and H. V. Gupta, Identification and evaluation of watershed models, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 29–47, AGU, Washington, D. C., 2003.

Young, P., Data-based mechanistic modeling and validation of rainfall-flow processes, in *Model Validation: Perspectives in Hydrological Science*, edited by M. G. Anderson and P. D. Bates, chap. 7, pp. x–x, John Wiley, Hoboken, N. J., 2001.

A. Sankarasubramanian, International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964-8000, USA. (sankar@iri.columbia.edu)

R. M. Vogel, WaterSHED Center, Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA. (richard.vogel@tufts.edu)