JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
APRIL AMERICAN WATER RESOURCES ASSOCIATION 2003

METHODOLOGIES FOR CALIBRATION AND PREDICTIVE ANALYSIS OF A WATERSHED MODEL[1]

John Doherty and John M. Johnston[2]

ABSTRACT: The use of a fitted parameter watershed model to address water quantity and quality management issues requires that it be calibrated under a wide range of hydrologic conditions. However, rarely does model calibration result in a unique parameter set. Parameter nonuniqueness can lead to predictive nonuniqueness. The extent of model predictive uncertainty should be investigated if management decisions are to be based on model projections. Using models built for four neighboring watersheds in the Neuse River Basin of North Carolina, the application of the automated parameter optimization software PEST in conjunction with the Hydrologic Simulation Program Fortran (HSPF) is demonstrated. Parameter nonuniqueness is illustrated, and a method is presented for calculating many different sets of parameters, all of which acceptably calibrate a watershed model. A regularization methodology is discussed in which models for similar watersheds can be calibrated simultaneously. Using this method, parameter differences between watershed models can be minimized while maintaining fit between model outputs and field observations. In recognition of the fact that parameter nonuniqueness and predictive uncertainty are inherent to the modeling process, PEST's nonlinear predictive analysis functionality is then used to explore the extent of model predictive uncertainty.
(KEY TERMS: mathematical modeling; HSPF; parameter estimation; PEST; model calibration; uncertainty; watershed management.)

Doherty, John and John M. Johnston, 2003. Methodologies for Calibration and Predictive Analysis of a Watershed Model. J. of the American Water Resources Association (JAWRA) 39(2):251-265.

INTRODUCTION

There are several advantages to be had in gaining computer assistance in model calibration. It is usually much easier to achieve a closer fit between model outputs and field measurements using a robust optimization package than can be achieved through manual calibration. However, there are other advantages in addition to efficiency. Principal among these is that most parameter optimization packages allow at least partial estimation of the degree of uncertainty associated with the values of optimized parameters. This, in turn, can lead to a better understanding of the margin of uncertainty that surrounds key predictions.

The use of computer based parameter estimation methods for the calibration of hydrologic models has received much attention in the literature. Effort has been devoted to the development of robust optimization algorithms for use in this context (Kuczera, 1983; Wang, 1991; Duan *et al.,* 1992; Sorooshian *et al.,* 1993; Sumner *et al.,* 1997). Other studies compare these methods (Gan and Biftu, 1996; Cooper *et al.,* 1997; Kuczera, 1997; Franchini *et al.,* 1998; Thyler *et al.,* 1999; Madsen *et al.,* 2002). Much attention has also been devoted to the best method of defining goodness-of-fit between model outcomes and field measurements (Kuczera, 1983; Gupta *et al.,* 1998; Liong *et al.,* 1998; Yapo *et al.,* 1998; Boyle *et al.,* 2000; Madsen, 2000; Madsen *et al.,* 2002). Further research has inquired into the uncertainty associated with predictions made by calibrated models (Beck, 1987; Kuczera and Parent, 1998; Kennedy and O'Hagan, 2001) and the level of complexity that is sustainable when these models are deployed in settings of limited data availability (Jakeman and Hornberger, 1993; Beven, 2000). There is a growing demand by environmental stakeholder groups and regulatory agencies that model predictive uncertainty analysis becomes a standard part of model deployment (NRC, 2001).

In the present study, we demonstrate computer assisted calibration and predictive uncertainty analysis using four hydrologic models deployed in neighboring watersheds. Variants of one particular calibration methodology are discussed in the context of exploring

[1]Paper No. 02004 of the *Journal of the American Water Resources Association.* **Discussions are open until October 1, 2003.**
[2]Respectively, Department of Environmental Engineering, University of Queensland, Australia; and USEPA/ORD/ NERL, Ecosystems Research Division, 960 College Station Road, Athens, Georgia 30605 (E-Mail/Johnston: johnston.johnm@epamail.epa.gov).

and partially overcoming some of the difficulties associated with parameter estimation and quantification of predictive uncertainty.

STUDY AREA

The Contentnea Creek basin, a Coastal Plain watershed, is located in the Neuse River basin in North Carolina, USA (Figure 1). Rainfall in the area averages 50 inches per year (Giese *et al.,* 1997). Mean annual maximum temperature is approximately 86°F, while mean minimum temperature is 50°F. The physiography is relatively uniform throughout the four modeled watersheds, with relatively low relief. The soils are well drained sands and sandy loams developed on sediments of marine origin.

A model was built for each of four nonoverlapping watersheds of the Contentnea Creek basin, these being the Contentnea Creek above Hookerton, Moccasin Creek, Nahunta Swamp, and Little Contentnea Creek. Areas of these watersheds are 311,924, 100,208, 52,815 and 57,692 acres, respectively. Each model was calibrated using daily streamflow records from U.S. Geological Survey (USGS) gauging stations (available online at http://waterdata.usgs.gov/nwis/). The models were built as part of a wider study dedicated to predicting alterations to water quality within the Contentnea Creek Basin as a result of increasing urbanization, changing farming practices, and climatic change (Johnston, 2001).

The primary land covers within all watersheds are forest, agriculture, and grassland/pasture; urban cover is about 3 percent (Sandra Bird, USEPA, personal communication, 2001). Land use classifications were taken from the Multi-Resolution Land Cover (MRLC) national land cover dataset (Vogelman *et al.,* 2001) with selected thematic map scenes being a composite of data acquired between 1990 and 1994. Of the four classes, the type responsible for the greatest differences in hydrological characteristics is the urban cover. Increases in impervious area result in flashy streamflow (i.e., peak flows immediately after rainfall and reduced baseflow in fair weather) and higher overland transport that increases constituent loading to streams.

METHODS

Hydrologic Model

Simulation of hydrologic processes within the four watersheds comprising the study area was undertaken using version 12 of the Hydrologic Simulation

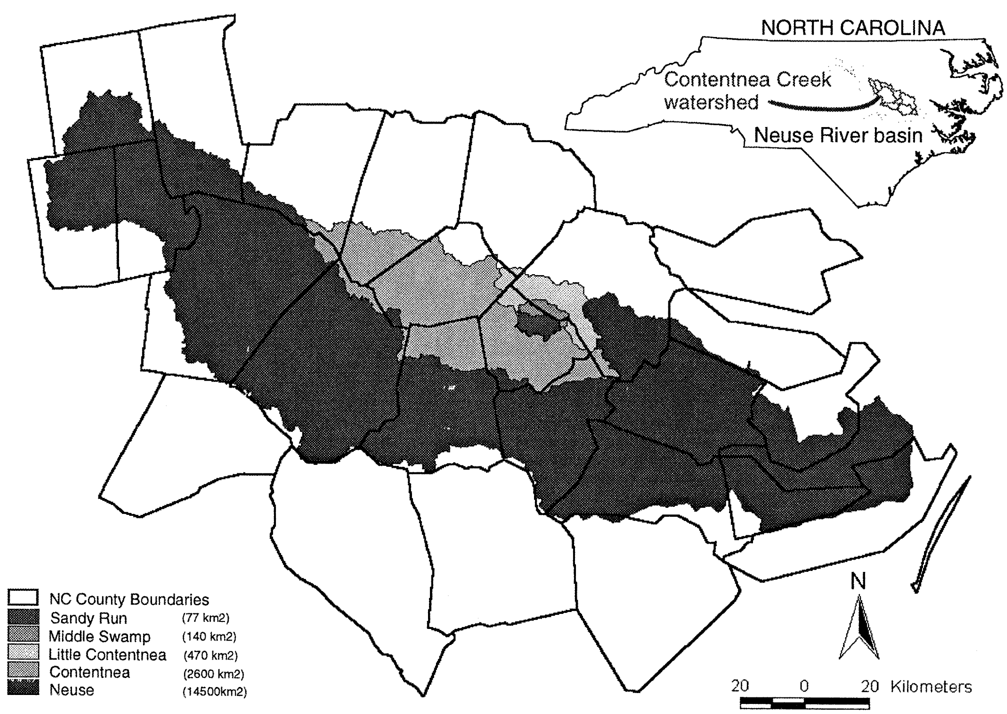

Figure 1. Contentnea Creek Watershed Study Area and Surroundings.

Program Fortran (HSPF) (see Bicknell *et al.,* 2001). Though a lumped parameter model, HSPF is considered moderately physically based. The parameters influencing water storage and flux are intended to be intuitive due to their correspondence to real world phenomena. Each subwatershed was simulated using four HSPF pervious land segments (PERLNDs), one IMPLND (impervious land segment), and a free flowing reach or mixed reservoir (RCHRES). The four PERLNDs represent the land use types mentioned above. The IMPLND was used for the simulation of urban impervious areas. The RCHRES simulates the flow of water in the tributary that drains each watershed.

Since the major hydrological difference between land use types is that between pervious and impervious land, initial model deployment was such that all four PERLNDs within each modeled watershed were assigned the same hydrologic parameters, except for the FOREST parameter that governs the amount of evapotranspiration taking place during winter. All parameters related to the physical dimensions of the system, for example land use areas, lengths of overland flow paths, and average slopes, were assigned in accordance with watershed geometry and topography.

Parameter Estimation Software

Model calibration was undertaken using PEST (Doherty, 2001a; Doherty, 2002) in conjunction with a suite of utility software written to support the use of PEST in the surface water modeling context (Doherty, 2001b). The principal member of this suite is TSPROC, a model independent time series processor optimized for use in the calibration context. PEST is a model independent parameter estimator with advanced predictive analysis and regularization features. Its model independence relies on the fact that it is able to communicate with a model through the latter's own input and output files, thus allowing easy calibration setup with an arbitrary model. Such a "model" can be encapsulated in a batch or script file if desired. Hence, model preprocessing and postprocessing software (such as TSPROC) can be used as part of the calibration process.

PEST implements a particularly robust variant of the Gauss-Marquardt-Levenberg method of parameter estimation. While this method requires that a continuous relationship exist between model parameters and model outputs, it can normally find the minimum in the objective function in fewer model runs than any other parameter estimation method. This is important when model run times are lengthy or when many parameters require estimation. In the present study model runs required one minute on a Pentium III 550 MHz machine. The Gauss-Marquardt-Levenberg method has been criticized for being too easily trapped in local objective function minima (Abbaspour *et al.,* 2001). In contrast, other algorithms such as the Shuffled Complex Evolution algorithm (Duan *et al.,* 1992, 1994) are much more likely to find the global objective function minimum. While this is an obvious advantage, the cost of guaranteeing that the global objective minimum is found is a much greater number of model runs than that required by the Gauss-Marquardt-Levenberg method. We circumvented the problem of local minima by formulating a calibration objective function that combines flows with various functions of flows, namely monthly volumes and exceedence times.

Outline of Approach Used in Present Investigation

The remainder of this paper briefly describes some of the methodologies used in calibrating the HSPF models for the four Contentnea Creek watersheds and in analyzing the uncertainty of predictions made by these models. The sequence is as follows:

1. First the basic parameter estimation methodology is applied to the calibration of a single watershed model.
2. Nonuniqueness of the calibrated parameters is examined, and other parameter sets are estimated that also calibrate the watershed model.
3. The ability of the calibrated model to make accurate predictions is examined by comparison of output time series with field measurements over a validation period.
4. Four watershed models are calibrated simultaneously, testing the hypothesis that model parameters pertaining to similar hydrologic units within adjacent watersheds have identical values.
5. The simultaneous model calibration process is repeated using a regularized calibration methodology in which parameter values for similar hydrologic units are allowed to differ, though only by the minimum amount required to achieve a user specified goodness-of-fit between model outcomes and field observations.
6. The effects of parameter nonuniqueness on the uncertainty of low flow model predictions are examined using PEST's nonlinear predictive analyzer.
7. The complexity of one of the watershed models is increased and the predictive analysis process repeated to demonstrate that the uncertainty range calculated using a complex model is more likely to encapsulate reality than that calculated using a simpler model.

Though described in the context of a specific set of watershed models, the methods described herein can be easily extended to other models and other watersheds.

CALIBRATION OF A SINGLE WATERSHED MODEL

Calibration Methodology

Calibration of each of the Contentnea Creek watershed models was undertaken by adjusting certain model parameters to obtain as good a match as possible between modeled and gauged flows over the period 1970 to 1985. Adjusted parameters, and their role in HSPF, are listed in Table 1. All of these parameters pertain to the HSPF PERLND module. As was mentioned above, the same values for these parameters were assigned to all four PERLNDs representing the four land use types within each watershed. The third column of Table 1 lists the initial values assigned prior to calibration adjustment, these values being considered reasonable for these watersheds (USEPA, 1999, 2000). The fourth column of Table 1 lists bounds on parameter values imposed throughout the calibration process. Note that HSPF employs a number of parameters in addition to those listed in Table 1. However, parameter estimation was limited to this group based on the sensitivity of these parameters in the calibration process. For most calibration exercises documented herein, the DEEPFR parameter was fixed at 0.1. This value was deemed reasonable as permanent loss of water to deep aquifers is considered unlikely to occur in any of the studied watersheds.

To reduce the nonlinearity of the parameter estimation problem and increase numerical stability, PEST estimated transformed interflow and groundwater recession parameters; these are related to the native HSPF parameters depicted in Table 1 by the following relationships:

$$\mathrm{IRCTRANS} = \mathrm{IRC}/(1\text{-}\mathrm{IRC}) \tag{1}$$

and

$$\mathrm{AGWRCTRANS} = \mathrm{AGWRC}/(1\text{-}\mathrm{AGWRC}) \tag{2}$$

These transformed parameters approach infinity as the native parameters approach 1.

All adjusted parameters were log transformed during parameter estimation to further increase the linearity of the problem and thereby reduce the chances of numerical instability.

In estimating values for model parameters, PEST minimized an objective function comprised of three components. These were the summed weighted squared differences between: (1) model generated and observed flows, (2) monthly volumes calculated on the basis of modeled and observed flows, and (3) exceedence times for various flow thresholds calculated on the basis of modeled and observed flows.

TABLE 1. HSPF Parameters, Their Functions, Initial Values, and Constraints Imposed During the Calibration Process.

Parameter Name	Parameter Function	Initial Value	Bounds*
LZSN	Lower zone nominal storage	5.0 in	2 to 15 in
UZSN	Upper zone nominal storage	0.5 in	0.01 to 2 in
INFILT	Related to the infiltration capacity of the soil	0.08 in/hour	0.001 to 0.5 in/hr
BASETP	The fraction of potential ET that can be sought from baseflow	0.1	0.01 to 0.2
AGWETP	Fraction of remaining potential ET that can be satisfied from active ground water storage	0.05	0.001 to 0.2
LZETP	Lower zone ET parameter – an index to the density of deep rooted vegetation	0.5	0.1 to 0.9
INTFW	Interflow inflow parameter	2.0	1.0 to 10.0
IRC	Interflow recession parameter	0.4/day	0.001 to 0.999/day
AGWRC	Ground water recession parameter	0.95/day	0.001 to 0.999/day
DEEPFR	Fraction of ground water inflow that goes to inactive ground water	0.1	Fixed

*Taken from USEPA, 2000.

Within the first of the above groups, weights assigned to individual flow observations were calculated using the formula:

$$w_i = c \times (1/f_i)^{1.5} \times (1 + \cos(2\pi d/365.25)/4) \tag{3}$$

where w_i is the weight assigned to flow observation i; f_i is the magnitude of flow observation i; c is a factor used to make the contribution to the objective function from each observation group similar in magnitude; and d is the day of the year (counting from January 1).

If observation weights are calculated as the reciprocals of the observations themselves, it can be shown that this is mathematically equivalent to calibration against the logs of observations. In calibrating a hydrologic model, such a strategy ensures that high flows do not dominate the parameter estimation process simply because of their large numerical value. The second factor in Equation (3) results in an even greater low flow weight than that provided through inverse magnitude weighting. This was done intentionally to focus the calibration process on low flows, hopefully increasing model performance during baseflow conditions. Decreased flows, particularly during dry, warm periods of the year, have the potential to impose risks on fish communities in many of the streams (Johnston, 2001).

The third factor in Equation (3) provides a means of partial discrimination against flows measured during the summer months when rainfall is likely to show a high degree of spatial heterogeneity. Discrepancies between hourly rainfall used to drive the model (taken from the nearest Class A weather station) and rainfall that actually occurred at the modeled watershed can complicate the calibration process. Since prevailing weather patterns in cooler months are generally more homogeneous over larger spatial extents, these periods are given more weight.

The use of a composite objective function comprised of a number of different criteria describing different aspects of the fit between model outputs and field data is now commonplace (see, for example, Madsen *et al.,* 2002; Madsen, 2000; Boyle *et al.,* 2000; Gupta *et al.,* 1998, 1999; and Yapo *et al.,* 1998). The use of multiple criteria must be accompanied by a suitable strategy for the selection of relative weights to apply when calculating the overall objective function. Weights were chosen such that no criterion was allowed to either dominate the objective function or to be dominated by another criterion. Other investigators, including those cited above, have adopted a more fluid weighting strategy, investigating the effects of different weights on parameters estimated through the optimization process. This leads to the concept of a "Pareto set" as a powerful mechanism for exploring parameter nonuniqueness. As will be discussed, parameter nonuniqueness is also explored in the present investigation within the context of fixed criterion weights. A fruitful direction for further research would be to extend the methodologies discussed herein to include parameter nonuniqueness induced by broadening of the choice of criterion weights.

For the initial parameter values listed in Table 1, the objective function for each of the watershed models was about 3 x 10^6. The contribution from each of the three observation groups (i.e., flows, monthly volumes and exceedence times) was roughly 1 x 10^6. For the Contentnea Creek model calibrated against flows recorded at Hookerton (henceforth referred to as the Hookerton model), PEST was able to reduce this objective function to 4.6 x 10^5 in about 100 model runs. Optimized parameter values are shown as "set 1" in Table 2. Graphical comparisons between modeled and measured flows through part of the calibration period, between modeled and observed monthly volumes over the entirety of the calibration period, and between modeled and observed exceedence times pertaining to the whole of the calibration period are shown in Figures 2a, 2b, and 2c. Note that the restriction of graphed flows in Figure 2a to only a part of the calibration period is done for the sake of clarity. Graphs over the remainder of the calibration period are similar. Note also that the flow axis is logarithmic in Figure 2 to afford a better comparison between model outputs and field measurements under both high and low flow conditions. Calibration results for the other watershed models are similar to those documented above for the Hookerton model.

A number of indices can be used to quantify the closeness of fit between modeled and observed time series (see Legates and McCabe,1999, for details). For the daily flows illustrated in Figure 2a, the coefficient of efficiency, modified coefficient of efficiency, index of agreement and modified index of agreement discussed by Legates and McCabe (with flows weighted according to Equation 3) are 0.77, 0.63, 0.95, and 0.81 respectively (note that the unmodified coefficient of efficiency was first introduced by Nash and Sutcliffe, 1970). For the time series of monthly volumes illustrated in Figure 2b, these indices are 0.87, 0.72, 0.86, and 0.97, respectively. In all future comparisons discussed in this study, we use only the modified coefficient of efficiency E_1 as recommended by Legates and McCabe. This is calculated using the formula:-

$$E_1 = 1 - \frac{\sum |S_i - O_i|}{\sum |O_i - \bar{O}|} \tag{4}$$

where S_i and O_i are the i^{th} terms of the simulated and observed time series, and $\bar{O}$ is the mean value of the

TABLE 2. Estimated Parameter Values. Parameter Set 1 was computed using unregularized parameter estimation. Parameter Sets 2 to 5 were computed using PEST's regularization functionality. Parameter Set 6 was computed through simultaneous calibration of the four watershed models.

Parameter Name	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
LZSN	2.0	2.0	2.0	2.0	2.0	2.0
UZSN	2.0	1.79	2.0	2.0	1.76	2.0
INFILT	0.0526	0.0615	0.0783	0.0340	0.0678	0.0687
BASETP	0.200	0.182	0.199	0.115	0.179	0.200
AGWETP	0.00108	0.0186	0.00232	0.0124	0.0247	0.0407
LZETP	0.50	0.50	0.20	0.72	0.50	0.50
INTFW	10.0	3.076	1.00	4.48	4.78	2.73
IRC	0.677	0.571	0.729	0.738	0.759	0.320
AGWRC	0.983	0.981	0.972	0.986	0.981	0.966
DEEPFR	0.1	0.1	0.1	0.1	0.1	0.1

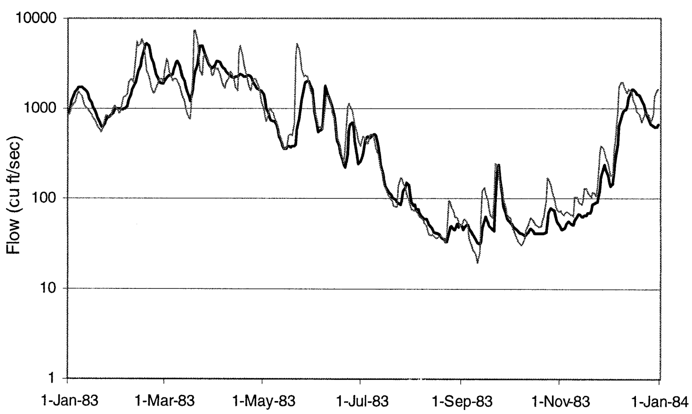

Figure 2a. Measured (bold line) and Modeled (light line) Hookerton Flows Over Part of the Calibration Period.

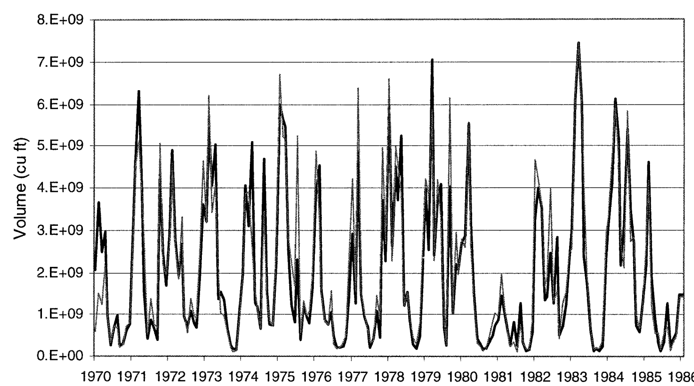

Figure 2b. Measured (bold line) and Modeled (light line) Monthly Volumes Over the Calibration Period.

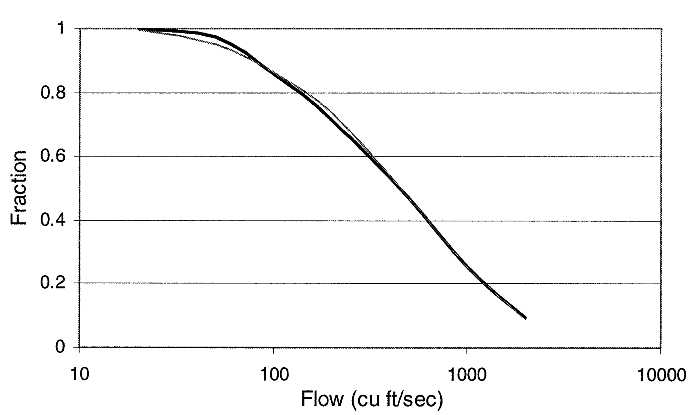

Figure 2c. Measured (bold line) and Modeled (light line) Flow Exceedence Fractions Over the Calibration Period.

observed time series. E_i can range from minus infinity to 1.0, with higher values indicating better agreement. If it exceeds zero, the model can be considered to be a better predictor of system behavior than the mean of the observed data.

As a by-product of the parameter estimation process, PEST calculates the composite scaled sensitivity (Hill, 1998) of each parameter. This is a measure of the sensitivity of all model outputs for which there are corresponding field measurements to each parameter. In the present instance, AGWRCTRANS and INFILT were the most sensitive parameters while INTFW and AGWETP were the least. Sensitivities spanned about an order of magnitude. It should be noted, however, that sensitivities can be highly dependent on parameter values for nonlinear models such as HSPF.

Parameter Nonuniqueness

It is possible to calibrate a rainfall/runoff model against a flow time series by adjusting only four or five parameters if a model is designed in such a way as to ensure maximum parameter sensitivity and minimum correlation between parameters (Jakeman and Hornberger, 1993). Correlation is the term used to describe the phenomenon whereby two or more parameters can be varied in harmony in such a way as to have virtually no effect on the calibration objective function. In the calibration process described herein, nine model parameters were adjusted to achieve an acceptable fit between model outcomes and measured flows, though adjustment for some parameters ceased when they hit their bounds. This suggests a degree of redundancy in the parameterization of the model, possibly involving correlation between the parameters in Table 1.

To determine if other sets of parameters could also calibrate the model, PEST was used in regularization mode. When run in this mode, the user supplies PEST with a default system condition expressed in terms of preferred values for parameters and/or preferred values for mathematical relationships between parameters. PEST is then used to calibrate the model to within a preferred model-to-measurement fit tolerance. This is defined through a limiting measurement objective function below which the model is deemed to be calibrated, while simultaneously minimizing a regularization objective function calculated on the basis of the misfit between optimized parameter values and their user-supplied default values or relationship values.

To find a number of different parameter sets that calibrate the Hookerton model, a number of different default system conditions were defined in terms of preferred values for the parameters listed in Table 1. In all cases these preferred values were within the bounds depicted in the fourth column of this table. A limiting measurement objective function of 5 x 10^5 was supplied for all PEST runs. This is slightly above that which it is possible to achieve without any regularization conditions being imposed, as established during the previous calibration exercise. It also allows a visually pleasing fit between measurements and model outcomes. The model was then recalibrated a number of different times using regularization to ensure that each calibrated parameter set departed to the smallest extent possible from the default parameter set supplied for that run. Four of the parameter sets determined in this way are listed as Sets 2 to 5 in Table 2. In all cases, the fit between model outcomes and observation data is commensurate with that depicted in Figures 2a, 2b, and 2c. Figure 3 compares modeled and measured flows over 1983 for Parameter Sets 2 to 5 from Table 2. The modified coefficients of efficiency (i.e., E_1) describing the fits between simulated and observed weighted daily flows for Parameter Sets 2 to 5 are 0.59, 0.62, 0.60, and 0.62; for monthly volumes they are 0.72, 0.73, 0.72, and 0.72. As is discussed above, E_i values corresponding to Parameter Set 1 are 0.63 and 0.72 for daily flows and monthly volumes, respectively.

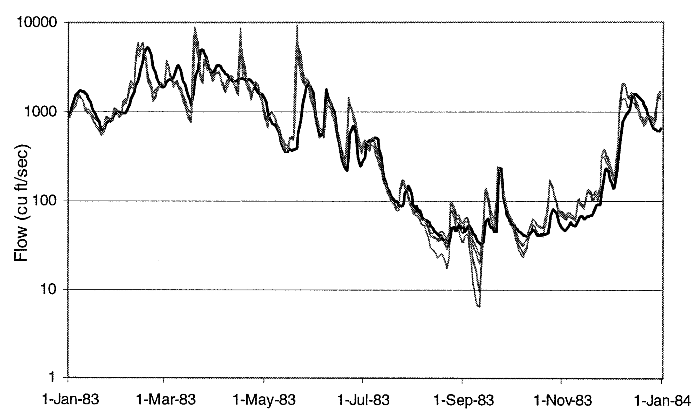

Figure 3. Measured (bold line) and Modeled (light lines) Hookerton Flows Over Part of the Calibration Period. Multiple parameter sets were estimated using PEST's regularization functionality.

The nonuniqueness of parameters estimated through the calibration process is readily apparent from these results. However, the extent of this nonuniqueness is not as dramatic as could have been achieved without the imposition of bounds on their values throughout the parameter estimation process. Lower zone nominal storage (LZSN) consistently encountered its lower bound, and upper zone nominal storage (UZSN) was usually at its upper limit. The imposition of these bounds left two less parameters to estimate, thereby reducing the amount of parameter redundancy. PEST's insistence on lowering LZSN to its two-inch lower bound and raising UZSN to its two-inch upper bound is noteworthy. Perhaps the tendency to go beyond their ranges indicates that they may play a role somewhat different from that which their names suggest.

Model Validation

The Contentnea Creek models were calibrated using flows recorded over the period 1970 to 1985. Flows recorded from 1986 to 1995 were used for validation. Figure 4a shows a comparison between observed and model generated flows for the Hookerton model over part of the validation period. Observed and model generated monthly volumes and observed and model generated exceedence fractions pertaining to the whole of the validation period are shown in Figures 4b and 4c. In these figures predictions made on the basis of Parameter Sets 1 to 5 listed in Table 2 are provided as grey lines. Bold lines represent measured flows, or quantities derived directly from them.

E_1 values describing the fits between modeled and measured weighted flows (with flow weights calculated using Equation 3) vary between 0.50 and 0.53 for Parameter Sets 1 to 5. For monthly volumes, E_1 varies between 0.70 and 0.72 – an acceptable fit in all cases. It thus appears that, even though the model calibration process resulted in a nonunique parameter set, predictions made by the calibrated model appear to be sensitive to the same *combinations* of parameters as those estimated through calibration. In general, this is more likely to occur when a model is used to make predictions that are of the same type as those against which it was calibrated. Where a model is used to make predictions of different types from those against which it was calibrated, or where model

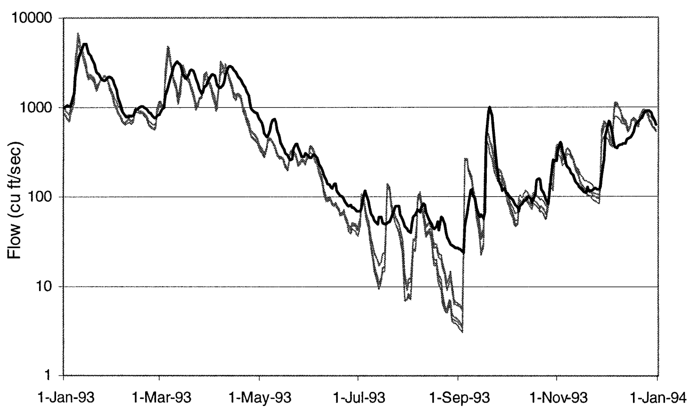

Figure 4a. Measured (bold line) and Modeled (light lines) Hookerton Flows Over Part of the Validation Period.

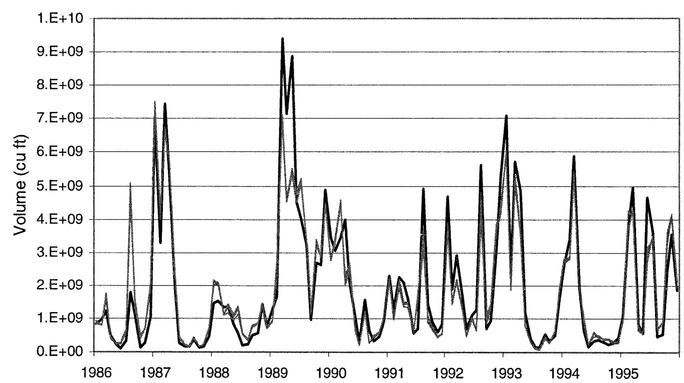

Figure 4b. Measured (bold line) and Modeled (light lines) Monthly Volumes Over the Validation Period.

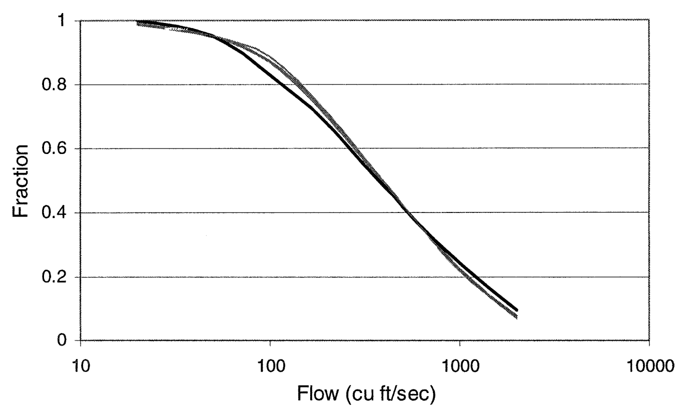

Figure 4c. Measured (bold line) and Modeled (light lines) Flow Exceedence Fractions Over the Validation Period.

inputs are significantly different under predictive conditions from those that prevailed under calibration conditions, opportunities arise for predictions to be sensitive to parameters, or parameter combinations, which are not well determined through the calibration process. In such circumstances, predictive uncertainty may be high. This occurs to some extent in the period around September 1, 1993, when flows are very low. This will be discussed further.

SIMULTANEOUS CALIBRATION OF MULTIPLE WATERSHED MODELS

Calibration Using Identical Parameter Values

From the foregoing discussion it is apparent that parameter uniqueness cannot be expected in the calibration of our watershed model. Calibration of the other three watershed models led to similar conclusions. In an attempt to reduce the degree of nonuniqueness in parameter estimates, extra information was introduced. The observation was made that topography, soil type, meteorology, and land use makeup are very similar across the four watersheds. It is therefore feasible that parameters assigned to PERLNDs representing identical land use types would be consistent across the watersheds.

In expanding Popper's (1959) exposition of the scientific method to the application of numerical simulation models in environmental management, Beck (1987) noted that environmental models can only be used to test hypotheses. A hypothesis can only be rejected, not accepted, on the basis of model usage. Using this principle, the hypothesis that PERLNDs representing identical land uses in all four watersheds can be assigned identical hydrologic parameter values was tested. The hypothesis can be rejected if the fit between field measurements and model outcomes is significantly inferior to that achieved through individual watershed model calibration. Strict application of statistical theory would dictate that an F-test be applied to flow, volume, and exceedence time residuals to test the hypothesis at various significance levels. However, visual inspection of overall goodness of fit is adequate in the present case.

A composite model was constructed through inclusion of the four individual watershed models in a single batch file in which the models are run sequentially. PEST was used to calibrate this composite model. Nine parameters were estimated (Table 1) with all land use PERLNDs in all four watershed models employing identical values for these parameters.

Parameter values estimated as an outcome of this process are those labeled as "Set 6" in Table 2. E_1 values for the Hookerton, Lucama, Nahunta, and Little Contentnea watersheds are 0.47, 0.44, 0.31, and 0.48 for weighted flows and 0.70, 0.58, 0.48, and 0.52 for weighted volumes. Figure 5 shows modeled and observed flows for the Hookerton model over 1983 (this being part of the 1970 to 1985 calibration period). Though visually pleasing over much of this period, the fit is unsatisfactory at low flows, where much of the attention of the present investigation is focused.

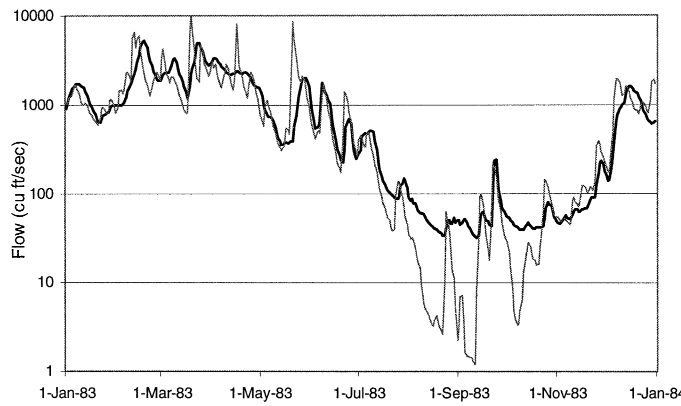

Figure 5. Measured (full line) and Modeled (light line) Hookerton Flows Over Part of the Calibration Period. Parameters were estimated through simultaneous calibration of all four watershed models.

Can we reject the hypothesis that parameters are the same for similar PERLNDs in the different watersheds? The answer is dependent on the intended use of the model. The fit between model outcomes and field measurements cannot be rejected as inadequate over the complete time series given the E_i values, and by visual inspection of Figure 5. However, our ability to make accurate predictions at low flows would probably be seriously degraded if we were to insist on using identical parameters for all watershed models. Nevertheless, the extent of model-to-measurement misfit may not be bad enough to reject the parameter set if used for other purposes, for example to parameterize an ungauged watershed in the same area for a preliminary analysis of its rainfall-runoff characteristics. For this latter application, the more watersheds with similar characteristics that are involved in the simultaneous calibration exercise, the more robust the parameter estimates are likely to be. The idea of prediction specific parameters that follows from this argument, together with the inherent nonuniqueness of parameters estimated through the calibration process, brings into question the idea that the model construction, calibration, and deployment process should ever yield a unique set of parameter values. Rather, model calibration should be viewed as a form of data interpretation. The manner in which data are most appropriately interpreted depends very much on the context in which that interpretation takes place as set by the environmental management issue that the model is being used to address.

Calibration Using Regularization

It was established that the ability of the Hookerton model to simulate low flows is compromised by requiring that its parameters be identical to those used by models deployed in neighboring watersheds. Nevertheless, hydrologic similarity with neighboring watersheds should not be ignored completely. There is information content in the hypothesis that variation of parameter values between adjacent watersheds should be minimal. In the present section, this information is introduced into the calibration process in a less restrictive form.

A parameter similarity condition can, in fact, be introduced to the parameter estimation process without compromising the level of model-to-measurement fit achieved by that process. Recall that in PEST's regularization mode the user sets the objective function below which the model is deemed to be calibrated. In attaining that objective function, PEST varies parameter values in such a way as to minimize the departure of these values from their preferred condition. However, attainment of the desired level of model-to-measurement fit is still PEST's primary goal.

In applying regularization to the simultaneous calibration of all four watershed models a preferred condition of cross watershed parameter equality was imposed. Because the limiting measurement objective function used in the parameter estimation process was set suitably low, an acceptable fit between model outputs and field measurements was obtained for all watersheds (E_1 values of 0.62, 0.44, 0.48, and 0.52 for weighted daily flows and 0.72, 0.58, 0.52, and 0.60 for monthly volumes). For the Hookerton model, the fits resemble those illustrated in Figures 2a, 2b, and 2c (see Figure 6). Estimated parameter values for all watersheds are listed in Table 3. Parameter differences between watersheds apparent in Table 3 exist because they *have to exist* to obtain the high level of fit illustrated in Figure 6. For comparison purposes, parameters estimated through independent watershed model calibration are shown italicized in brackets in Table 3. Interwatershed variation is obviously much greater for these parameters.

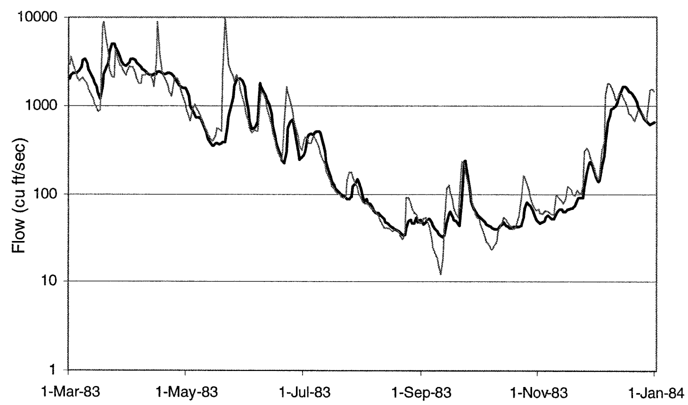

Figure 6. Measured (full line) and Modeled (light line) Hookerton Flows Over Part of the Calibration Period. Model parameters were estimated through simultaneous calibration of all watersheds using regularization.

TABLE 3. Parameters Estimated by PEST Through Simultaneous Watershed Model Calibration Using Its Regularization Functionality. Parameters estimated through independent model calibration are shown in parentheses.

Parameter Name	Contentnea at Hookerton	Moccasin at Lucama	Nahunta Swamp	Little Contentnea
LZSN	2.29 (2.00)	2.01 (2.00)	2.58 (3.244)	2.00 (2.00)
UZSN	2.00 (2.00)	2.00 (2.00)	2.00 (2.00)	1.55 (1.93)
INFILT	0.0533 (0.0526)	0.0317 (0.0194)	0.0706 (0.117)	0.0276 (0.00518)
BASETP	0.163 (0.20)	0.182 (0.118)	0.157 (0.20)	0.166 (0.114)
AGWETP	0.0201 (0.00108)	0.0269 (0.0493)	0.0222 (0.00358)	0.0268 (0.00814)
LZETP	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)
INTFW	1.21 (10.0)	1.00 (1.00)	1.17 (1.406)	1.31 (3.253)
IRC	0.533 (0.670)	0.506 (0.794)	0.512 (0.220)	0.499 (0.799)
AGWRC	0.988 (0.984)	0.967 (0.980)	0.976 (0.967)	0.942 (0.956)
DEEPFR	0.1 (fixed)	0.1 (fixed)	0.1 (fixed)	0.1 (fixed)

ANALYZING THE UNCERTAINTY OF MODEL PREDICTIONS

Predictive Analysis

Predictions made under conditions that differ from those which prevail during calibration are likely to be more uncertain than those made under similar conditions. This phenomenon is exemplified in the Hookerton model's failure to accurately predict the low flows that occurred over the few days centered on September 1, 1993. Figure 4a shows that predictions made using all of the estimated parameter sets undercalculate flow over this time, a particularly worrying phenomenon, as the calibration process attempted to optimize the model's ability to predict such low flows. Also apparent from Figure 4a is the uncertainty surrounding flow predictions made over this time.

Multiple recalibration using regularization in conjunction with different default parameter values is one way of exploring model predictive uncertainty. Using this methodology, a model can be calibrated many times with a different parameter set estimated each time. Predictions can then be made using all estimated parameter sets. However, a far more efficient way to explore predictive uncertainty is to first identify a specific prediction whose uncertainty requires exploration and then to find a parameter set that maximizes/minimizes that prediction while maintaining the model in a calibrated state (as defined by an upper objective function limit below which the model is deemed to be calibrated). This was accomplished using PEST's nonlinear predictive analysis functionality. Like nonlinear parameter estimation, nonlinear predictive analysis as implemented in PEST is an iterative procedure involving many model runs. Though a numerically intensive process, it is by far the most effective means available for exploration of the uncertainty surrounding a specific prediction. The algorithm underpinning PEST's predictive analysis functionality is based on the theory presented by Vecchia and Cooley (1987) (see also Doherty, 2001a, for further details).

Total flow volume over the August 29 to September 3, 1993, period was identified as the specific model prediction to maximize and minimize while maintaining the model within calibration bounds relative to measured flows, volumes, and exceedence times spanning the period 1970 to 1985. The limiting objective function defining calibration was the same as that used earlier in exploring the role of regularization in estimating parameter sets that deviate minimally from a set of user supplied preferred values. Figure 7a shows model calculated flows over 1993 based on parameter sets for which the key prediction was maximized and minimized. Figure 7b shows model-to-measurement fits for these same two parameter sets over part of the calibration period. In the former figure the low flow period for which flows were maximized and minimized is circled for ease of recognition. In each of these figures the dashed, light colored curve represents the output of the minimization model, whereas the full, light colored curve represents the output of the maximization model. Measured flows are represented using a bold line.

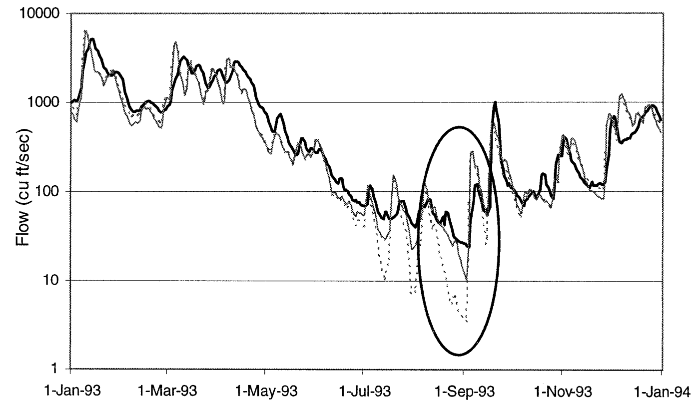

Figure 7a. Model Generated and Measured Hookerton Flows Over Part of the Validation Period. Model parameters were estimated using PEST's predictive analysis functionality with flow minimized (dashed light line) and maximized (full light line) over the highlighted period.

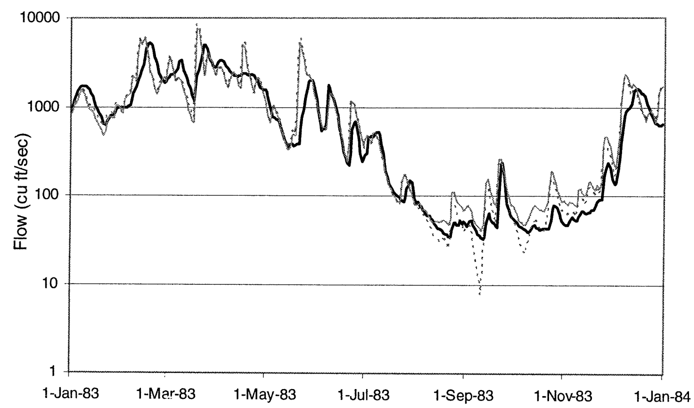

Figure 7b. Model Generated (light lines) and Measured (bold line) Hookerton Flows Over Part of the Calibration Period. Model parameters were estimated using PEST's predictive analysis functionality.

The range of uncertainty accompanying the prediction of flows near September 1, 1993, is apparent from an inspection of Figure 7a. As Figure 7b demonstrates, both the model used for prediction maximization and that used for prediction minimization fit measured flows well under calibration conditions (E_1 values for weighted daily flows are 0.62 in both cases, while those for monthly volumes are 0.71 and 0.72 for prediction minimization/maximization). Calibrated parameters for the minimization and maximization models are Sets 7 and 8, respectively, in Table 4.

Model Complexity

It is unfortunate that even with the predicted flow maximized over the six-day period of interest, the model generated flow is less than the observed. This can be construed as an inability on the part of the model to replicate all of the temporal fine detail of the system's behavior. (Whether it is actually necessary for a model to replicate such fine detail depends on the uses to which the model will be put.)

In general, if a model is to simulate system fine detail, it must be constructed with an appropriate level of complexity. The introduction of complexity to a model is generally accompanied by the introduction of extra parameters. It has already been demonstrated that, even though the Hookerton model can be adequately calibrated with the current number of parameters, these parameters cannot be estimated uniquely. Hence, the introduction of more parameters is likely to increase the extent of parameter nonuniqueness, even if it increases the model's ability to replicate system detail.

To introduce more complexity, the DEEPFR parameter that had previously been fixed was allowed to vary. PEST was then used to adjust this parameter, along with the parameters that it had already been adjusting, to minimize and maximize flow at Hookerton over the August 29 to September 3, 1993, period while, once again, maintaining the model in a calibrated state over the period 1970 to 1985. Figure 8a shows flows over 1993 predicted using the maximization and minimization parameter sets, while Figure 8b shows flows during 1983 (part of the calibration period) calculated on the basis of these two parameter sets. Estimated parameters for minimization and maximization of flow are listed as Set 9 and Set 10, respectively, in Table 4. E_1 values for weighted daily flows and monthly volumes over the calibration period are 0.62 and 0.72, respectively, for both prediction minimization and maximization.

An inspection of Figure 8a reveals that measured flows over the six-day period spanning August 29 to September 3, 1993, are now at the margin of predictive uncertainty of the model, and that these margins have increased in both directions as a result of the introduction of the extra complexity. (Note that these margins could have been expanded even more by increasing the limiting calibration objective function applied during the predictive analysis process.) This illustrates an important aspect of model usage in environmental simulation. In general, while it is true that system fine detail can be replicated only if the necessary complexity is introduced to a model, the heightened level of parameter correlation and insensitivity that results from the addition of that complexity often results in higher levels of uncertainty surrounding predictions of system fine detail. In other words, because a model *can* simulate complex processes, this does not guarantee that it *will* simulate them with precision. If the appropriate level of complexity is included in a model, all that can be guaranteed is

Table 4. Estimated Parameter Values. All parameter sets were estimated using PEST's predictive analysis functionality. Parameter Sets 7 and 8 were computed with DEEPFR fixed. Parameter Sets 7 and 9 minimize flow volume over selected prediction interval while Parameter Sets 8 and 10 maximize it.

Parameter Name	Set 7	Set 8	Set 9	Set 10
LZSN	2.0	2.0	2.0	2.0
UZSN	1.90	2.0	1.58	1.91
INFILT	0.0675	0.030	.0871	0.029
BASETP	0.20	0.20	0.20	0.20
AGWETP	0.0169	0.001	0.022	0.001
LZETP	0.50	0.50	0.50	0.50
INTFW	4.73	10.0	5.44	10.00
IRC	0.587	0.671	0.65	0.833
AGWRC	0.980	0.990	0.979	0.995
DEEPFR	0.1 (fixed)	0.1 (fixed)	0.166	0.262

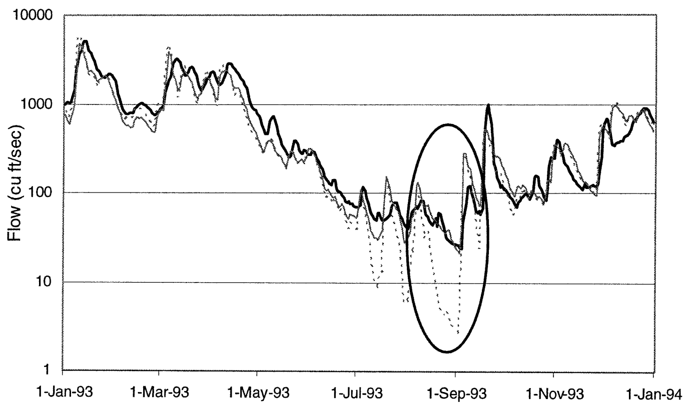

Figure 8a. Model Generated (light lines) and Measured (bold line) Hookerton Flows Over Part of the Validation Period. Model parameters were estimated using PEST's predictive analysis functionality with DEEPFR adjustable.

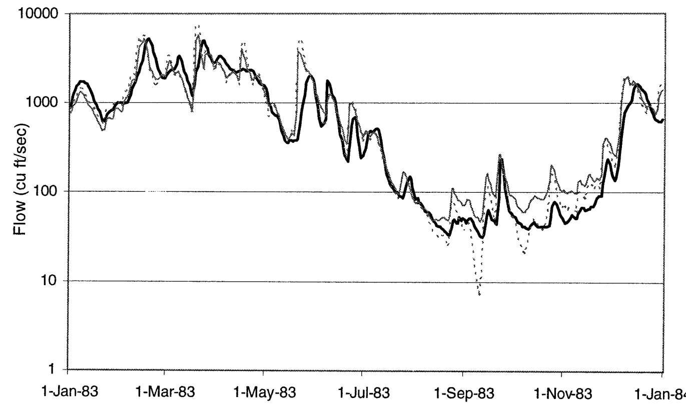

Figure 8b. Model Generated (light lines) and Measured (bold line) Hookerton Flows Over Part of the Calibration Period. Model parameters were estimated using PEST's predictive analysis functionality with DEEPFR adjustable.

that true system behavior will lie somewhere within the uncertainty limits of predictions made by that model. The introduction of complexity to a model thus endows the modeler with the ability to know the boundaries of future system behavior, but not necessarily the details. The need for predictive uncertainty analysis in conjunction with model deployment -- especially if a model is deployed to investigate system fine detail – is thus paramount.

It is worth mentioning that theory is available that relates the limiting calibration objective function applied during the predictive analysis process to a specific prediction probability (Vecchia and Cooley, 1987). However, application of this theory is difficult in the present case due to the fact that the selection of weights for different components of the objective function is still a somewhat arbitrary process; furthermore, this theory does not account for the fact that individual criterion weights could be varied considerably and still provide a suitable basis for model calibration.

CONCLUSIONS

Though focused on a particular environmental management problem, the purpose of this paper has been to demonstrate new methodologies for environmental data processing based on the use of numerical simulation models in conjunction with sophisticated parameter estimation and predictive analysis software. A number of methodologies have been demonstrated, the application of which was intended to extract the maximum information content from particular environmental datasets in ways that allow most beneficial processing of that data to address particular management issues.

Calibration of a model such as HSPF can rarely be achieved without at least some degree of nonuniqueness in the estimated parameters. This leads to uncertainties in model predictions, especially predictions at the extremes of system behavior. By using PEST's regularization functionality it is possible to explore the extent of parameter nonuniqueness by calculating different sets of model parameters, all of which calibrate the model to within a tolerance specified by the user. Model predictions can then be made using all of these parameter sets, allowing the uncertainties associated with these predictions to be gauged.

By undertaking simultaneous calibration of a number of different models constructed for similar watersheds, the hypothesis that hydrologic parameters are likely to be similar for these different watersheds can be tested in the parameter estimation process. The use of regularized parameter estimation allows recognition of this condition of preferred parameter equality while allowing each individual model to remain calibrated to within a user specified tolerance.

Nonlinear predictive uncertainty analysis provides another mechanism for exploration of the uncertainty associated with key model predictions. The user is able to maximize or minimize a particular model output while ensuring that the model remains calibrated to within a specified tolerance.

A brief analysis of the role of model complexity was undertaken. It was demonstrated that the addition of complexity to a model does not guarantee that model predictions will be more accurate than those made by a simpler model. However, the inclusion of complexity in a model may allow better estimation of the degree of uncertainty associated with key model predictions, and may provide a stronger guarantee that true system behavior will lie within the margins of predictive uncertainty.

DISCLAIMER

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

John Doherty wrote much of the time series analysis software that underpins the work documented above while employed as a Visiting Research Scientist at the University of Idaho, Idaho Falls. He wishes to acknowledge the funding and resources made available to him from that institution that allowed the development of this software to take place.

The assistance provided by Mike Cyterski (USEPA) in construction of the models upon which this work was based is gratefully acknowledged. Stephen Kraemer (USEPA), Brenda Rashleigh (USEPA), and Sandy Bird (USEPA) are also thanked. The comments of Gary Shenk (USEPA, Chesapeake Bay Program) and two anonymous reviewers improved the manuscript.

Both PEST and the utility software used to carry out the analyses documented herein are available from the first author, or from the EPA CEAM web site: http://www.epa.gov/ceampubl/sitemap.htm.

LITERATURE CITED

Abbaspour, K. C., R. Schulin, and M. Th. van Genuchten, 2001. Estimating Unsaturated Soil Hydraulic Properties Using Ant Colony Optimization. Adv. in Water Resour. 24:827-841.

Beck, B., 1987. Water Quality Modeling: A Review of the Analysis of Uncertainty. Water Resour. Res. 23 (8):1393-1442.

Beven, K. J., 2000. Uniqueness of Place and Process Representations in Hydrological Modeling. Hydrology and Earth Systems Sciences 4(2):203-213.

Bicknell, B. R., J. C. Imhoff, J. L., Kittle, T. H. Jobes, and A. S. Donigian, 2001. HSPF User's Manual. Aqua Terra Consultants, Mountain View, California.

Boyle, D. P., H. V. Gupta, and S. Sorooshian, 2000. Toward Improved Calibration of Hydrologic Models: Combining the Strengths of Manual and Automatic Methods. Water Resour. Res. 36(12):3663-3674.

Cooper, V. A., V. T. V. Nguyen, and J. A. Nicell, 1997. Evaluation of Global Optimization Methods for Conceptual Rainfall-Runoff Model Calibration. Water Sci. Tech. 36(5):53-60.

Doherty, J., 2001a. PEST-ASP User's Manual. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., 2001b. PEST Surface Water Utilities User's Manual. Watermark Numerical Computing, Brisbane, Australia, and University of Idaho, Idaho Falls, Idaho.

Doherty, J., 2002. Addendum to PEST Manual for Version 6 of PEST. Watermark Numerical Computing, Brisbane, Australia.

Duan, Q. S., S. Sorooshian, and V. K. Gupta, 1992. Effective and Efficient Global Optimization for Conceptual Rainfall Runoff Models. Water Resour. Res. 28 (4):1015-1031.

Duan, Q. S., S. Sorooshian, and V. K. Gupta, 1994. Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models. J. Hydrol. 158:265-284.

Franchini, M., G. Galeati, and S. Berra, 1998. Global Optimization Techniques for the Calibration of Conceptual Rainfall-Runoff Models. Hydrol. Sci. J. 43(3):443-458.

Gan, T. Y. and G. F. Biftu, 1996. Automatic Calibration of Conceptual Rainfall-Runoff Models: Optimization Algorithms, Catchment Conditions and Model Structure. Water Resour. Res. 32(12):3513-3524.

Giese, G. L., J. L. Eimers, and R. W. Coble, 1997. Simulation of Ground-Water Flow in the Coastal Plain Aquifer System of North Carolina. *In:* Regional Aquifer-System Analysis – Northern Atlantic Coastal Plain. U.S. Geological Survey Professional Paper 1404–M, 142 pp.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, 1998. Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information. Water Resour. Res. 34(4):751-763.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, 1999. Status of Automatic Calibration for Hydrologic Models: Comparison With Multilevel Expert Calibration. J. Hydrol. Eng. 4(2):135-143.

Hill, M. C., 1998. Methods and Guidelines for Effective Model Calibration. U.S. Geological Survey Water Resources Investigations Report 98-4005.

Jakeman, A. J. and G. M. Hornberger, 1993. How Much Complexity is Warranted in a Rainfall-Runoff Model? Water Resour. Res. 29(8):2637-2649.

Johnston, J. M., 2001. A Scientific and Technological Framework for Evaluating Risk in Ecological Risk Assessments. *In:* Modeling of Environmental Chemical Exposure and Risk, J. B. H. J. Linders (Editor). Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 133-150.

Kennedy, M. C. and A. O'Hagan, 2001. Bayesian Calibration of Computer Models. J. R. Statist. Soc. B. 63 (3):425-464.

Kuczera, G., 1983. Improved Parameter Inference in Catchment Models. 1. Evaluating Parameter Uncertainty. Water Resour. Res. 19(5):1151-1172.

Kuczera, G., 1997. Efficient Subspace Probabilistic Parameter Optimization for Catchment Models. Water Resour. Res. 33(1):17-185.

Kuczera, G. and E. Parent, 1998. Monte Carlo Assessment of Parameter Uncertainty in Conceptual Catchment Models: The Metropolis Algorithm. J. Hydrol. 211:69-85.

Legates, D. R. and G. J. McCabe, 1999. Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. Water Resour. Res. 35(1):233-241.

Liong, S. Y., S. T. Khu, and W. T. Chan, 1998. Derivation of Pareto Front With Accelerated Convergence Genetic Algorithm, ACGA. *In:* Hydroinformatics 1998, V. Babovic and L. C. Larsen (Editors). Balkema, Rotterdam, The Netherlands, pp. 889-896.

Madsen, H., 2000. Automatic Calibration of a Conceptual Rainfall-Runoff Model Using Multiple Objectives. J. Hydrol. 235:276-896.

Madsen, H., G. Wilson, and H. C. Ammentorp, 2002. Comparison of Different Automated Strategies for Calibration of Rainfall-Runoff Models. J. Hydrol. 261:48-59.

Nash, J. E. and J. V. Sutcliffe, 1970. River Flow Forecasting Through Conceptual Models. I. A Discussion of Principles. J. Hydrol. 10:282-290.

NRC (National Research Council), 2001. Assessing the TMDL Approach to Water Quality Management. National Academy Press, Washington, D.C., 109 pp.

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

Sumner, N. R., P. M. Flemming, and B. C. Bates, 1997. Calibration of a Modified SFB Model for Twenty-Five Australian Catchments Using Simulated Annealing. J. Hydrol. 197:166-188.

Sorooshian, S., Q. Duan, and V. K. Gupta, 1993. Calibration of Rainfall-Runoff Models: Application of Global Optimization to the Sacramento Soil Moisture Accounting Model. Water Resour. Res. 29(4):1185-1194.

Thyler, M., G. Kuczera, and B. C. Bates, 1999. Probabilistic Optimization for Conceptual Rainfall-Runoff Models: A Comparison of the Shuffled Complex Evolution and Simulated Annealing Algorithms. Water Resour. Res. 35(3):767-773.

USEPA (U.S. Environmental Protection Agency), 1999. HSPFParm: An Interactive Database of HSPF Model Parameters, Version 1.0. EPA-823-R-99-004, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

USEPA (U.S. Environmental Protection Agency), 2000. BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF. EPA-823-R-00-012, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Vecchia, A. V. and R. L. Cooley, 1987. Simultaneous Confidence and Prediction Intervals for Nonlinear Regression Models With Application to a Groundwater Flow Model. Water Resour. Res. 23(7):1237-1250.

Vogelman, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel, 2001. Completion of the 1990's National Land Cover Data Set for the Conterminous United States. Photogrammetric Engineering and Remote Sensing 67:650-662.

Wang, Q. J., 1991. The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resour. Res. 27(9):2467-2471.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1998. Multi-Objective Global Optimization for Hydrologic Models. J. of Hydrol. 204:83-97.