WATER RESOURCES RESEARCH, VOL. 43, W01402, doi:10.1029/2006WR004883, 2007

Comparison of hydrologic calibration of HSPF using automatic and manual methods

Sang Min Kim,[1] Brian L. Benham,[1] Kevin M. Brannan,[1] Rebecca W. Zeckoski,[1]
and John Doherty[2,3]

Received 11 January 2006; revised 8 August 2006; accepted 6 September 2006; published 5 January 2007.

[1] The automatic calibration software Parameter Estimation (PEST) was used in the hydrologic calibration of Hydrological Simulation Program–Fortran (HSPF), and the results were compared with a manual calibration assisted by the Expert System for the Calibration of HSPF (HSPEXP). In this study, multiobjective functions based on the HSPEXP model performance criteria were developed for use in PEST, which allowed for the comparison of the calibration results of the two methods. The calibrated results of both methods were compared in terms of HSPEXP model performance criteria, goodness-of-fit measures (R^2, E, and $RMSE$), and base flow index. The automatic calibration results satisfied most of the HSPEXP model performance criteria and performed better with respect to R^2, E, $RMSE$, and base flow index than manual calibration results. The results of the comparison with the manual calibration suggest that the automatic method using PEST may be a suitable alternative to manual method assisted by HSPEXP for calibration of hydrologic parameters for HSPF. However, further research of the weights used in the objective functions is necessary to provide guidance when applying PEST to surface water modeling.

Citation: Kim, S. M., B. L. Benham, K. M. Brannan, R. W. Zeckoski, and J. Doherty (2007), Comparison of hydrologic calibration of HSPF using automatic and manual methods, *Water Resour. Res.*, *43*, W01402, doi:10.1029/2006WR004883.

1. Introduction

[2] The Hydrological Simulation Program–Fortran (HSPF) [*Bicknell et al.*, 2001] is a comprehensive watershed model that is in wide use but that requires calibration. Currently, HSPF is one of the primary models used to develop total maximum daily loads (TMDLs). The TMDL program, which is mandated by the Clean Water Act (33 U.S.C. §§ 1251–1387), is a watershed management process that integrates watershed planning with water quality assessment and protection. The U.S. Environmental Protection Agency (U.S. EPA) estimates that public and private costs associated with TMDL development over the next 15 years will be in excess of $1 billion [*U.S. EPA*, 2001]. Given this large investment of both public and private funds, research is needed into methods to improve the use and application of HSPF to ensure these funds are invested wisely and result in measurable water quality improvement.

[3] Because HSPF requires calibration, improving the efficiency and accuracy of the calibration process has the potential to reduce TMDL development costs and increase the usefulness of the model. Many HSPF users perform the hydrologic calibration of the model manually using the decision-support software Expert System for the Calibration of HSPF (HSPEXP) [*Lumb et al.*, 1994]. This paper explores the use of the automated Parameter Estimation software (PEST) [*Doherty*, 2004] as a possible alternative to manual calibration aided by HSPEXP.

[4] Manual calibration is time consuming and often tedious [*Madsen*, 2000]. Furthermore, because of the subjectivity involved and the lack of uniqueness of the calibrated parameter set, it is difficult to explicitly assess the confidence of the model simulations and to maintain consistency across users. As a result, the expertise acquired by one individual through extensive hands-on training and experience with a specific model is not easily transferred to another person (or another model). Nonetheless, for an experienced hydrologist it is possible to obtain a good calibration using the manual approach.

[5] Because of the time-consuming and difficult nature of manual calibration, research has been directed to development of more effective and efficient automatic calibration procedures [*Madsen et al.*, 2002]. With automatic calibration, parameters are adjusted according to a specified search scheme in order to minimize an objective function. Compared with manual calibration, automatic calibration is fast and the confidence of the model simulations can be explicitly stated. In addition, a robust optimization package is typically easier to use than manual calibration and may result in better agreement between model outputs and observed data. However, many surface water hydrologists do not consider automatic calibration acceptable for various reasons. As a result, automatic calibration has not entered into widespread use for surface water hydrologic and water quality models [*Boyle et al.*, 2000].

---

[1] Biological Systems Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

[2] School of Engineering, University of Queensland, Brisbane, Queensland, Australia.

[3] Now at Watermark Numerical Computing, Corinda, Queensland, Australia.

0043-1397/07/2006WR004883

Table 1. HSPF Model Parameters Adjusted During Hydrologic Calibration and Their Initial Values

Parameter	Definition	Initial Value	Recommended Range
LZSN	lower zone nominal storage, mm	218.4	50.8–381
UZSN	upper zone nominal storage, mm	17.53	0.25–50.8
INFILT	index to infiltration capacity, mm/h	1.78–15.49	0.025–16
BASETP	fraction of potential ET that can be sought from base flow	0–0.1	0–0.2
AGWETP	fraction of remaining potential ET that can be satisfied from active groundwater storage	0–0.001	0–0.2
LZETP	lower zone ET parameter, an index to the density of deep rooted vegetation	0.1–0.8	0.1–0.9
INTFW	interflow inflow parameter	1.0–2.0	1.0–10.0
IRC	interflow recession parameter, per day	0.6	0.001–0.999
AGWRC	groundwater recession parameter, per day	0.99	0.001–0.999
DEEPFR	fraction of groundwater inflow that goes to inactive groundwater	0.0	0–0.2
CEPSC	interception storage capacity, mm	1.27–6.35	0.25–10.16

[6] Many past applications of automatic calibration procedures focused on using a single overall objective function to measure performance of the calibrated model. Calibration based on a single performance measure, however, is often inadequate to properly measure the simulation of all the important characteristics of a hydrologic system. This aspect contributes to skepticism regarding automatic calibration [*Yapo et al.*, 1998; *Madsen*, 2000].

[7] Recently, multiple-objective functions have been used for automatic calibration of surface water models. Multiple-objective functions include a number of different criteria describing different aspects of fit between model outputs and observed data, and their use is now commonplace [*Madsen et al.*, 2002; *Boyle et al.*, 2000; *Gupta et al.*, 1998, 1999; *Yapo et al.*, 1998]. The U.S. National Weather Service typically uses as many as nine different criteria to measure the performance of the Sacramento Soil Moisture Accounting model (SAC-SMA) during a multistage, semi-automated calibration procedure [*Brazil*, 1988]. PEST is a very flexible, model-independent program that can utilize multiple-objective functions in the calibration process.

[8] PEST has been widely used in the field of groundwater modeling, but there have been very few applications of PEST to surface water models. Recently, PEST has been applied to HSPF calibration [*Doherty and Johnston*, 2003], using daily flow, monthly flow, and exceedance time as subcomponents of the objective function and a coefficient of efficiency to evaluate goodness-of-fit. This research builds on their work by creating multiple-objective functions based on the HSPEXP model performance criteria and then assessing the adequacy of PEST as an alternative to HSPEXP for calibrating HSPF.

2. Methods

[9] For this study, HSPF was applied to a watershed in Virginia. Multiple-objective functions, based on the HSPEXP model performance criteria, were developed for use in PEST. Two independent calibrated models were developed, one using automated calibration with PEST and the other performing a manually assisted calibration with HSPEXP. The HSPEXP model performance criteria, coefficient of determination (R^2), coefficient of efficiency (E), root-mean-square error ($RMSE$), and base flow index were used to compare the resulting calibrated models.

2.1. Description of HSPF

[10] HSPF is a comprehensive, continuous, lumped parameter, watershed-scale model that simulates the movement of water, sediment, and a wide range of water quality constituents on pervious and impervious surfaces, in soil profiles, and within streams and well-mixed reservoirs [*Bicknell et al.*, 2001]. HSPF comprises three main modules: PERLND, IMPLND, and RCHRES. The PERLND module represents pervious land, the IMPLND module represents impervious surface area where little or no infiltration occurs, and the RCHRES module represents the stream reaches and reservoirs in a watershed. In this research, only the hydrologic processes were simulated; sediment and water quality were not considered. The initial values for the HSPF hydrologic parameters were estimated based on guidance from *U.S. EPA* [2000]. Initial parameter values and recommended ranges for the parameters adjusted during both the PEST and HSPEXP hydrology calibrations are listed in Table 1.

2.2. Calibration of HSPF

2.2.1. Manual Calibration

[11] When performing a manual HSPF calibration, the HSPEXP decision-support software guides the modeler using expert advice compiled from experienced modelers. An acceptable HSPF calibration is achieved when the percent errors between the simulated and observed data for HSPEXP model performance variables listed in Table 2 are less than or equal to the limits set by the modeler. All seven HSPEXP model performance variables were used in the manual calibration; however, only six of the seven variables were considered in this study, as Virginia does not require the inclusion of the low-flow recession variable in hydrology calibration for TMDLs [*Lawson*, 2003]. HSPEXP provides default values for model performance criteria (Table 2). The modeler may increase or decrease specific criteria based on modeling objectives, quality of the observed data, or other justifiable reasons; for this study, the default criteria were used. The HSPEXP decision-support software provides calibration guidance, suggesting parameter adjustments that sequentially address total volume, then low flows, storm flows, and finally seasonal flows. HSPEXP calculates percent errors of the model performance variables for each calibration run and provides guidance based on the values of those errors relative to predefined criteria. HSPEXP version 2.4 was used in this study. Actual parameter adjustments were made considering both the advice of HSPEXP and the modeler's prior experience using HSPF. The parameters LZSN, LZETP, INFILT, and AGWRC were independently estimated and calibrated for each land use during the calibration. The

Table 2. HSPEXP Model Performance Variables and Criteria for Hydrologic Calibration of HSPF[a]

HSPEXP Model Performance Variable	HSPEXP Model Performance Criteria, % Error
Total volume	±10
Fifty-percent lowest flows	±10
Ten-percent highest flows	±15
Storm peaks	±15
Seasonal volume error	±10
Summer storm volume error	±15

[a] From *Benham et al.* [2005].

parameters UZSN, LZETP, and CEPSC were input on a monthly basis and varied across land use. The remaining parameters (BASETP, AGWETP, INTFW, IRC, and DEEPFR) were constant throughout the year and were not varied across land use. Manual calibration ceased when all the HSPEXP model performance criteria listed in Table 2 were met.

2.2.2. Automatic Calibration

[12] Automatic calibration was conducted using the Parameter Estimation (PEST) software. PEST is a model-independent parameter optimization program that minimizes one or more user-specified objective functions. PEST implements a particularly robust variant of the Gauss-Marquardt-Levenberg [*Marquardt*, 1963] method of nonlinear 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 [*Doherty*, 2004]. Because PEST is able to communicate with a model through the model's own input and output files, it can be used to estimate parameters for many existing computer simulation models, whether or not the user has access to model source codes; in fact the "model" may comprise one or more executables run sequentially through a batch or script file. A principal component of PEST, TSPROC, is a model-independent time series processor. TSPROC acts as a preprocessor to prepare the PEST control and instruction files as well as a model postprocessor during PEST optimization runs [*Doherty*, 2004]. Through the use of TSPROC it is a relatively simple matter to construct a multicomponent objective function involving a large number of observations. TSPROC undertakes the model-run postprocessing that is necessary for formulation of the objective function; it also automates generation of output that PEST requires for minimization of the objective function (or functions).

[13] As a by-product of its implementation of the Gauss-Marquardt-Levenberg (GML) method in minimization of the user-specified objective function, PEST is able to provide linear-based approximations of the uncertainties pertaining to parameters that it estimates, and of the degree of correlation between them, this being a reflection of the inherent nonuniqueness of the inverse problem that it is asked to solve.

[14] During the automatic calibration process using PEST, the same parameters were adjusted as with the manual calibration (Table 1). Six subobjective functions that were based on HSPEXP model performance criteria were combined to form a single composite objective function. Weights were assigned to each subobjective function to ensure that the contributions of each to the multiple-objective functions were almost equal [*Doherty and Johnston*, 2003]. This was done in order to ensure that no subobjective function dominated the inversion process. Thus the information content of the original calibration data set that each component of the objective function attempted to "distill" was made as visible as possible to the inversion process undertaken by PEST. It is acknowledged that the process of assigning weights to components of a multicomponent objective function can be subjective. In an attempt to remove such subjectivity, the concept of "Pareto optimality," which determines the effect of different weighting strategies on optimized parameters, was applied as a part of the calibration process by *Doherty and Johnston* [2003]. The main focus of our research was the possible benefits of using the HSPEXP performance variables when applying PEST to the hydrologic calibration of HSPF. The subjectivity of weight assignment is an important topic and considered in future research. The multiple-objective functions can be stated as the optimization problem: min(with respect to θ) $F(\theta) = \{f_1(\theta), \cdots, f_m(\theta)\}$, where $f_i(\theta)$ are the subobjective functions that are simultaneously minimized with respect to the parameter set (θ) of the model. The subobjective functions developed from the HSPEXP model performance criteria are listed in Table 3, where Q is daily flow, EX is the fraction of time that stream flow equals or exceeds a specific flow rate, N_{st} is the total number of selected storm events in the calibration period, P is peak flow, N_{ss} is the number of summer and winter months, n_j is the number of time steps in each month j, m_j is the number of time steps in each storm event j, θ is the set of model parameters to be calibrated, and w is a weighting function.

2.2.3. Goodness-of-Fit Measures

[15] Some common goodness-of-fit statistics, in addition to the HSPEXP model performance criteria, were used to assess HSPF performance for each calibration method. These goodness-of-fit statistics were not used directly in either the automatic or manual calibration processes, but were used after the calibrations were completed as an additional method of comparison. The coefficient of determination (R^2), coefficient of efficiency (E), and root-mean-square error ($RMSE$) have been widely used to evaluate the goodness-of-fit of hydrologic models. The coefficient of determination (R^2) is the square of the Pearson's product-moment correlation coefficient and describes the proportion of the total variance in the observed data that can be explained by the model. It ranges from 0.0 to 1.0, with higher values indicating better agreement, and is given by

$$R^2 = \left\{ \frac{\sum_{i=1}^{N} \left(Q_{obs,i} - \bar{Q}_{obs}\right)\left(Q_{sim,i} - \bar{Q}_{sim}\right)}{\left[\sum_{i=1}^{N} \left(Q_{obs,i} - \bar{Q}_{obs}\right)^2\right]^{0.5} \left[\sum_{i=1}^{N} \left(Q_{sim,i} - \bar{Q}_{sim}\right)^2\right]^{0.5}} \right\}^2, \quad (1)$$

where the over bar denotes the mean for the entire time period of the evaluation. *Nash and Sutcliffe* [1970] defined a coefficient of efficiency (E) that ranges from minus infinity to 1.0, with higher values indicating better agreement. If E is greater than zero, the model is considered to be a better

Table 3. The Formula and Its Description of Multiobjective Functions Developed for Hydrologic Calibration of HSPF Using PEST

Description	Formula
Squared error of daily flows	$f_1(\theta) = \sum_{i=1}^{N} [(Q_{obs,i} - Q_{sim,i}(\theta)) \cdot w_{1,i}]^2$
Squared error of 50% lowest flows exceedance	$f_2(\theta) = [(EX_{obs,50\%\ \mathrm{lowest\ flow}} - EX_{sim,50\%\ \mathrm{lowest\ flow}}(\theta)) \cdot w_{2,i}]^2$
Squared error of 10% highest flows exceedance	$f_3(\theta) = [(EX_{obs,10\%\ \mathrm{highest\ flow}} - EX_{sim,10\%\ \mathrm{highest\ flow}}(\theta)) \cdot w_{3,i}]^2$
Squared error of storm peaks	$f_4(\theta) = \sum_{i=1}^{N_{st}} [(P_{obs,i} - P_{sim,i}(\theta)) \cdot w_{4,i}]^2$
Squared error of seasonal volume	$f_5(\theta) = \sum_{j=1}^{N_{ss}} [(\sum_{i=1}^{n_j} Q_{obs,i} - \sum_{i=1}^{n_j} Q_{sim,i}(\theta)) \cdot w_{5,i}]^2$
Squared error of storm volume	$f_6(\theta) = \sum_{j=1}^{N_{st}} [(\sum_{i=1}^{m_j} Q_{obs,i} - \sum_{i=1}^{m_j} Q_{sim,i}(\theta)) \cdot w_{6,i}]^2$

predictor of system behavior than the mean of the observed data. The coefficient of efficiency (E) is defined as

$$E = 1.0 - \frac{\sum_{i=1}^{N} \left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{N} \left(Q_{obs,i} - \bar{Q}_{obs}\right)^2}. \tag{2}$$

Root-mean-square error ($RMSE$) is an absolute error measure quantifying the error in terms of the units of the variable, and is given by

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} \left(Q_{obs,i} - Q_{sim,i}\right)^2\right]^{0.5}. \tag{3}$$

[16] Finally, the base flow index, which is the ratio of base flow to total flow, was used to compare the calibration results. Base flow for the observed and the simulated flow time series for each calibration method was estimated using the HYSEP software [*Sloto and Crouse*, 1996].

3. Study Watershed

[17] The North River watershed was used for comparison of the two calibration methods. The North River watershed is 972.8 km^2 in size and located in Rockingham and Augusta counties, Virginia (Figure 1). North River is a tributary of the South Fork of the Shenandoah River (U.S. Geological Survey (USGS) Hydrologic Unit Code 02070005), which flows into the Potomac River and eventually discharges into the Chesapeake Bay.

3.1. Meteorological and Hydrological Data

[18] The primary meteorological data source used for model input was the National Climatic Data Center's (NCDC) Cooperative Weather Station at Dale Enterprise (COOPID 442208), Rockingham County, Virginia, located within the North River watershed. The long-term record summary (1 August 1948 to 31 March 2004) for the Dale Enterprise station shows average annual precipitation to be 903.5 mm. Average annual daily temperature is 11.8°C. The highest average daily temperature of 23.2°C occurs in July while the lowest average daily temperature of 0.2°C occurs in January [*Southeast Regional Climate Center*, 2004]. Data from the Staunton Sewage Plant station (COOPID 448062) located in Augusta County, Virginia, were used to supplement data missing from the Dale Enterprise record. A weather input data file for the HSPF model was created for the period January 1985 through December 2003 using the Watershed Data Management Utility (WDMUtil) [*Hummel et al.*, 2005]. Raw data required for creating the weather data file included hourly precipitation, average daily temperatures (maximum, minimum, and dew point), average daily wind speed, total daily solar radiation, and percent sun. Historical daily mean stream flow data for the USGS gauge station "North River near Burketown" (01622000; 78°54′50″W, 38°20′25″N; see Figure 1) were obtained from the USGS Web site [*USGS*, 2005].

3.2. Model Input Data Descriptions

[19] Following the guidance outlined in U.S. EPA Technical Note 6, topographic, stream network, soil, and land use data required to develop inputs for HSPF were obtained and processed using ArcGIS 8.3 [*Environmental Systems*

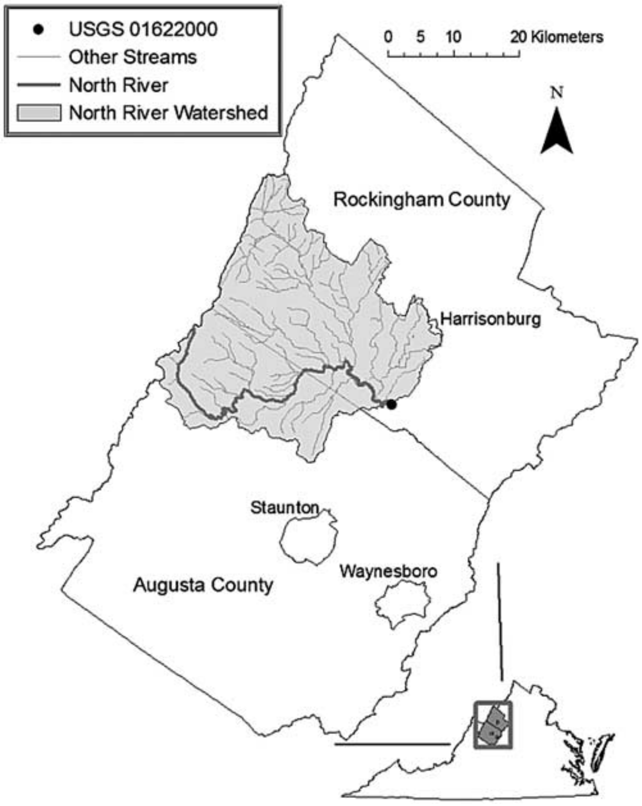
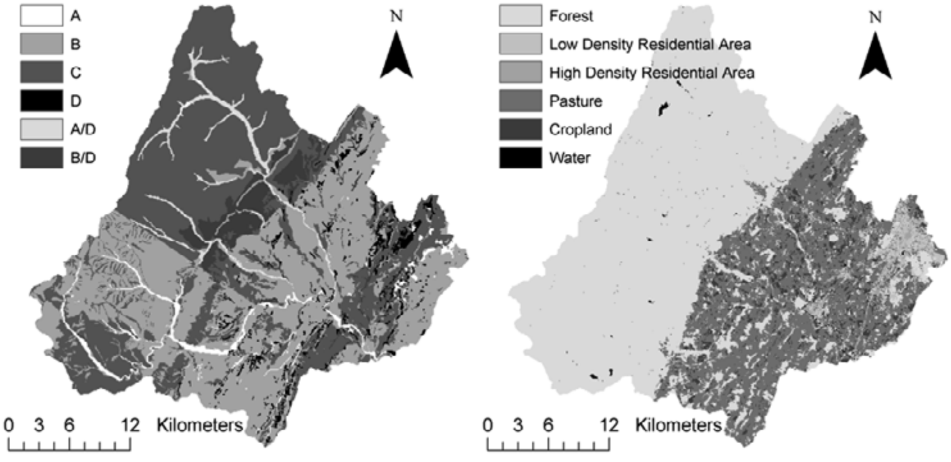


Figure 1. Location of the North River watershed.



(a) (b)

Figure 2. Hydrologic soil groups and land use of the North River watershed.

Research Institute, 2003] to estimate the initial parameter values for HSPF. Topographic data were obtained from 1:24,000 scale U.S. Geological Survey (USGS) digital elevation models (DEMs). The detailed stream network was obtained from the U.S. Census Topologically Integrated Geographic Encoding and Referencing (TIGER) data.

[20] Soils information was obtained from the Soil Survey Geographic (SSURGO) database provided by the Natural Resources Conservation Service (NRCS). The dominant hydrologic soil groups in the watershed were B and C, at 41.9% and 44.6%, respectively. Hydrologic soil groups B and C are characterized by moderate and low infiltration rates, respectively. Figure 2a shows the delineated hydrologic soil groups for the North River watershed.

[21] The land use data were obtained from the 1992 National Land Cover Dataset (NLCD), which is a 21-class land cover classification scheme derived from Landsat Thematic Mapper satellite data. The North River watershed included 15 of these classes. The 15 land use classes in North River watershed were grouped into five major categories: forest, low-density residential, high-density residential, pasture, and cropland, based on similarities in hydrologic features. The main land use category in the North River watershed is forest (62%), followed by pasture (29%). The remaining land uses include cropland (5%), low-density residential (3%), and high-density residential (1%). Most of the forest areas are located in the northwestern part of the watershed (Figure 2b).

[22] Hydraulic function tables, FTABLEs, were created to represent hydraulic characteristics of the reaches from topographic information using the method described by *Staley et al.* [2006]. Using this method, bankfull entries for the FTABLEs were generated from NRCS Regional Hydraulic Geometry Curves for the Piedmont Upland Region [*NRCS*, 2004]; floodplain entries for the FTABLEs were generated from basic geometric analysis of the DEM and Manning's equation.

3.3. Selection of Calibration/Validation Periods

[23] A 19-year period from January 1985 through December 2003 was examined with the goal of identifying two time periods with representative ranges of moderately high and moderately low rates of precipitation and runoff. These two time periods would be used for calibration and validation. Figure 3 shows the daily rainfall and runoff from 1985 to 2003 at the outlet of North River watershed. For this period, the annual mean rainfall was 932.2 mm and runoff was 384.9 mm. The rainfall and runoff were highest in 1996 with the second highest amounts in 2003 (Figure 4). The annual mean runoff ratio was 0.43 and ranged from 0.25 in 1999 to 0.57 in 1996. Mean monthly rainfall and the resulting monthly runoff were highly variable with the

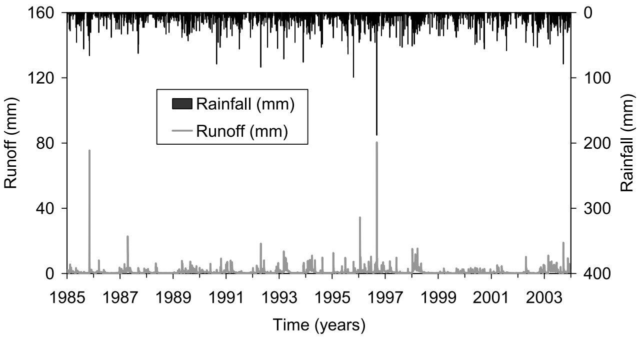


Figure 3. North River watershed daily rainfall and runoff from 1985 to 2003.

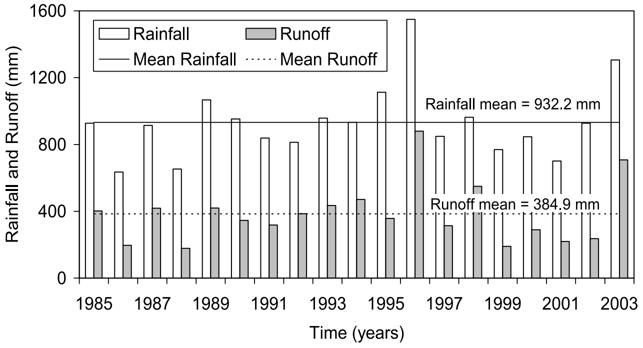


Figure 4. North River watershed annual rainfall and runoff from 1985 to 2003.

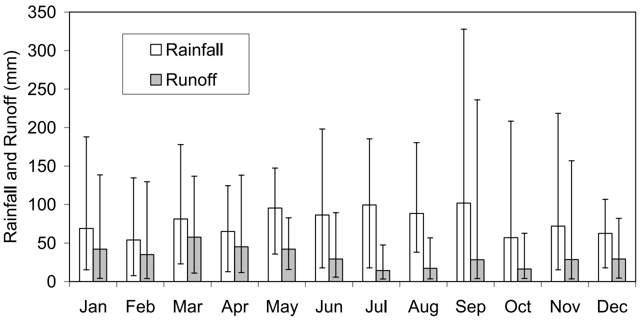


Figure 5. North River watershed mean monthly rainfall and runoff. (Error bars represent data ranges.)

greatest variability in September and least variability in December. The lowest runoff observed in most years occurred in summer, although the rainfall was relatively high during most summers (Figure 5). As a result, the average runoff ratios in July and August were below 0.2, the lowest seen over the year. At the other extreme, the runoff ratio was highest in March at 0.71, followed by a second peak in April at 0.69.

[24] The late 1990s marked an extended period of below-average rainfall and runoff conditions (Figure 3) that was difficult for the model to accurately represent. Precipitation and runoff in 1996 and 2003 were well above normal. Furthermore, the land use data were for 1992 conditions. Therefore a period including 1992 was desirable. The period of 1985–1994 represented a good mixture of below average, above average, and average precipitation and runoff conditions; allowed a sufficient continuous period of record for both calibration and validation; and included the year represented by the land use data. The selected calibration period was 1 September 1985 to 31 August 1990, while the validation period was 1 September 1990 to 31 December 1994.

[25] HSPEXP requires the user to select storm events within the simulation period in order to compute storm-related HSPEXP model performance criteria. Storm events were selected based on the recommendations in BASINS Technical Note 5 [*U.S. EPA*, 1999]. The starting date for each storm was the day precipitation began, and the ending date was the day when observed flow returned to prestorm (or almost prestorm) conditions. Selected storms for both the calibration and validation periods included both high and moderate to low peaks to represent the full range of storm types. The 23 storms selected for the automatic and manual calibration are listed in Table 4. Selected storm peaks for the calibration period ranged from 1.7 to 849.5 m^3/s; the average of all 23 peaks was 68.4 m^3/s. The 13 storms selected for the validation period are listed in Table 5. Selected storm peaks for the validation period ranged from 6.9 to 207.0 m^3/s; the average of all 13 peaks was 50.4 m^3/s.

4. Results

4.1. Comparison of Calibration Results

[26] The model parameters listed in Table 1 were adjusted using both the manual and automatic calibration methods for the period from 1 September 1985 to 31 August 1990. Table 6 shows the final calibrated parameter values for both methods. There is general agreement among the automatic and manual calibrated parameter sets. However, the automatic method estimated lower values for UZSN and a higher value for INTFW compared with the manual method. For the manual calibration, the parameters INFILT, LZETP, and UZSN were adjusted separately for the individual land uses. For the automatic method, INFILT, LZETP, and UZSN values were held constant across the land uses. We initially allowed these parameters to vary across land use; however, PEST did not provide better results and thus we decided to hold them constant across land uses to reduce computation time and improve numerical stability. For the PEST calibration of monthly parameters, the authors used a two-parameter sine curve instead of directly calibrating 12 monthly parameters. The two parameters of the sine curve were adjusted by PEST. Using this approach, the bounds of monthly parameters could not be strictly limited. This is the reason that the lower bounds of LZETP were slightly exceeded for the automatic calibration. In future research, methods of parameter regularization will be used to address these issues. To reduce the nonlinearity of the parameter estimation problem and increase numerical stability, AGWRC and IRC were transformed; these are related to the native HSPF parameters depicted in Table 1 by the following relationships:

$$IRCTRANS = IRC/(1 - IRC) \tag{4}$$

$$AGWRCTRANS = AGWRC/(1 - AGWRC). \tag{5}$$

Table 4. Selected Storm Events for the Calibration Period

Storm Starting Date	Storm Ending Date	Storm Duration, days	Storm Peak, m^3/s	Storm Starting Date	Storm Ending Date	Storm Duration, days	Storm Peak, m^3/s
30 Oct 1985	9 Nov 1985	10	849.5	13 Jan 1989	21 Jan 1989	8	9.4
11 Dec 1985	13 Dec 1985	2	11.9	4 Mar 1989	11 Mar 1989	7	19.9
11 Mar 1986	23 Mar 1986	12	91.5	30 Apr 1989	3 May 1989	3	88.1
19 May 1986	25 May 1986	6	26.2	16 Jun 1989	26 Jun 1989	10	36.2
8 Jul 1986	11 Jul 1986	3	2.4	25 Jul 1989	28 Jul 1989	3	20.7
9 Aug 1986	12 Aug 1986	3	1.7	14 Sep 1989	19 Sep 1989	5	56.4
21 Dec 1986	30 Dec 1986	9	41.1	15 Oct 1989	25 Oct 1989	10	33.1
23 Apr 1987	28 Apr 1987	5	45.3	15 Nov 1989	20 Nov 1989	5	51.8
6 Sep 1987	15 Sep 1987	9	37.7	29 Dec 1989	5 Jan 1990	7	40.2
24 Dec 1987	4 Jan 1988	11	26.1	8 Feb 1990	14 Feb 1990	6	22.2
17 Jan 1988	27 Jan 1988	10	32.3	21 Aug 1990	29 Aug 1990	8	26.9
28 Aug 1988	31 Aug 1988	3	3.8				
Average						6.7	68.4

Table 5. Selected Storm Events for the Validation Period

Storm Starting Date	Storm Ending Date	Storm Duration, days	Storm Peak, m^3/s	Storm Starting Date	Storm Ending Date	Storm Duration, days	Storm Peak, m^3/s
10 Jan 1991	21 Jan 1991	11	80.7	10 Nov 1992	19 Nov 1992	9	23.6
22 Mar 1991	29 Mar 1991	7	78.2	20 Jan 1993	28 Jan 1993	8	17.3
22 Jun 1991	28 Jun 1991	6	9.6	14 Apr 1993	19 Apr 1993	5	72.8
24 Jul 1991	4 Aug 1991	11	17.4	5 Sep 1993	10 Sep 1993	5	6.9
19 Apr 1992	29 Apr 1992	10	207.0	5 May 1994	12 May 1994	7	92.3
21 Jul 1992	25 Jul 1992	4	12.2	19 Jul 1994	23 Jul 1994	4	20.9
3 Sep 1992	8 Sep 1992	5	16.4				
Average						7.1	50.4

[27] Graphical comparisons of the observed and simulated daily flows, monthly flows, and exceedance time for the calibration period are shown in Figures 6, 7, and 8. A logarithmic axis is used in Figure 6 to permit better comparison between observed and modeled discharge under low-flow conditions. Overall, the simulated daily flows resulting from the automatic and manual calibration methods agreed very well with the observed flows, with the exception of an extreme storm event on 5 November 1985. The runoff from the 5 November 1985 storm was 74.8 mm, 3.3 times greater than the next greatest storm event. Neither calibration method was able to capture this extreme event. There was generally good agreement between the simulated monthly flows, Figure 7. The simulation results using the PEST calibration parameter set performed better for the middle-low to middle-high flow range with respect to the exceedance time, Figure 8. Both calibrated parameter sets over-predicted low flows and under-predicted midrange flows.

[28] The results of the HSPF simulation using both the automatic and manual calibration parameter sets were further compared using the HSPEXP model performance criteria and the conventional goodness-of-fit statistics R^2, E, and $RMSE$. Table 7 summarizes model performance for both calibration methods for the study watershed during the 5-year calibration period. The manual calibration method satisfies all the HSPEXP model performance criteria, while the automatic calibration method slightly violates the criteria for the lowest 50% of flows by 1.15% and the criteria for average storm peaks by 4.52%. However, the PEST calibration results showed better agreement in total runoff than the HSPEXP calibration results, with an error of −2.93% compared with HSPEXP's −9.59%. For the conventional goodness-of-fit statistics, the automatic calibration method produced better results for all three statistics when compared with the manual calibration method.

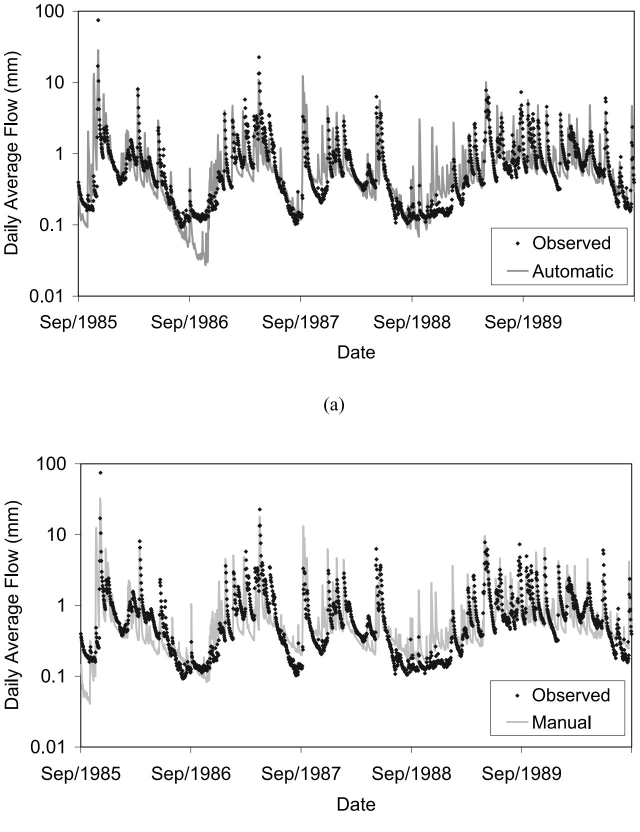
Table 6. Estimated Parameter Values by Automatic and Manual Calibration Methods

Parameter Name	Final Value		Bounds
	Automatic Method	Manual Method	
LZSN	117.3 mm	127 mm	50.8–381 mm
UZSN	0.30–0.31 mm[a]	1.78–12.7 mm[a]	0.25–50.8 mm
INFILT	2.76 mm/h	0.25–12.7 mm/h	0.025–16 mm/h
BASETP	0.017	0	0–0.2
AGWETP	0.002	0	0–0.2
LZETP	0.0037–0.7245[a]	0.1–0.6[a]	0.1–0.9
INTFW	10	3	1.0–10.0
IRC	0.523 per day	0.6 per day	0.001–0.999
AGWRC	0.984 per day	0.98–0.99 per day	0.001–0.999
DEEPFR	0.1 (fixed)	0.06	0–0.2
CEPSC	1.27–3.81 mm[a] (fixed)	1.27–3.81 mm[a]	0.25–10.16 mm

[a] Varies on a monthly basis.

4.2. Comparison of Validation Results

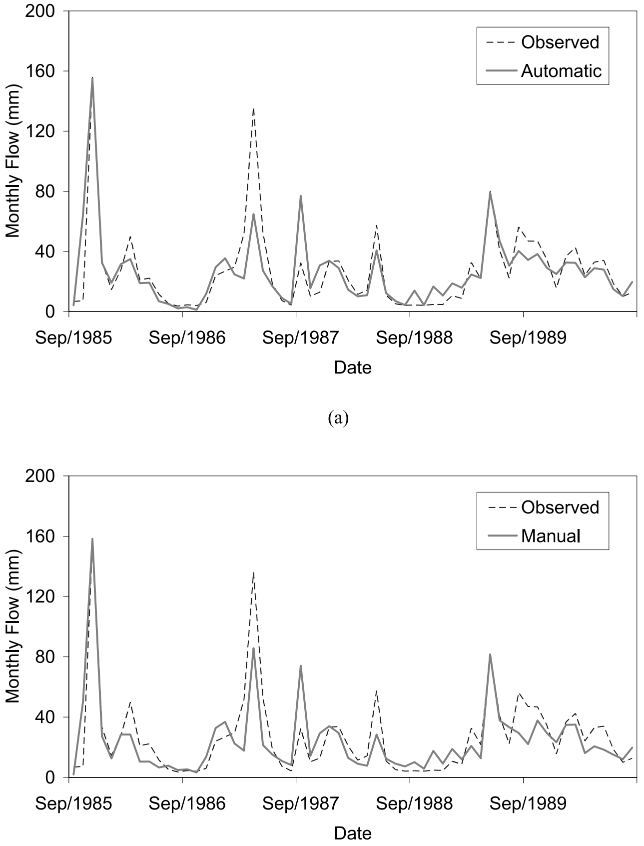
[29] The calibrated parameters were validated for the period from 1 September 1990 to 31 December 1994. Figure 9 shows the daily flows during the validation period. Overall, the simulated daily flows agreed well with the observed flows for both automatic and manual parameter sets. There were some discrepancies in low flows (below 0.2 mm) in the automatic parameter set. The graphical comparison of monthly volumes is shown in Figure 10. Both automatic and manual parameter sets showed a similar



(a)

(b)

Figure 6. Observed and simulated daily flows over the calibration period using (a) automatic and (b) manual methods.



(a)

(b)

Figure 7. Observed and simulated monthly flows over the calibration period using (a) automatic and (b) manual methods.

trend and agreed relatively well with the observed data during the validation period. With respect to exceedance time (Figure 11), the simulated flows from the automatic calibration agreed better with observed flows in the middle-low to middle-high flow range during the validation period. Overall, both parameter sets over-predicted low flows and under-predicted mid-flows; this was also observed during the calibration period.

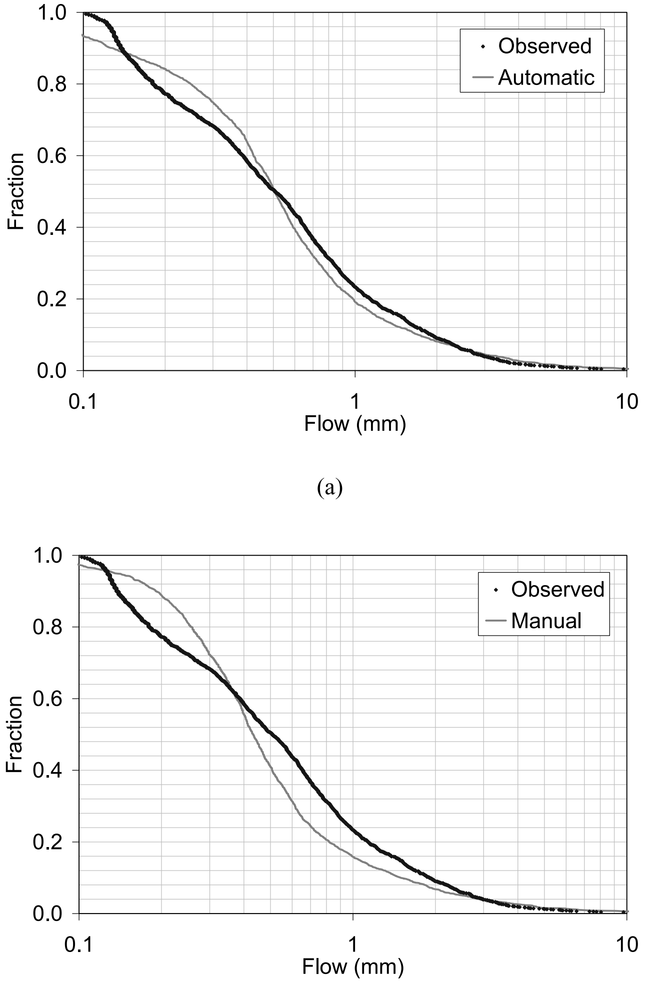
[30] The validation period results were also compared using HSPEXP model performance criteria and conventional goodness-of-fit statistics. Table 8 shows the summary of the statistics for the 4.5-year validation period. The simulated flows using the automatic calibration parameter set satisfied all the HSPEXP model performance criteria except the error in summer storm volumes. However, the simulated flows produced by automatic calibration parameter set agreed better with total observed runoff than those produced by the manual parameter set (see Table 8). As was the case for the calibration period, the automatic calibration parameter set produced better results in all three conventional goodness-of-fit statistics (see Table 8).

4.3. Comparison of Base Flow index

[31] Base flow estimates for the calibration and validation periods are shown in Table 9. Observed flows were partitioned into surface flow and base flow using HYSEP [*Sloto and Crouse*, 1996]. When the observed flow data were evaluated using HYSEP, the base flow indices for the calibration and validation periods were both 0.55. For the calibration period, the base flow indices were 0.49 for the automatic calibration and 0.47 for the manual calibration. The base flow indices were 0.44 (automatic method) and 0.39 (manual method) for the validation period. The automatic calibrated parameter set provided a slightly better match to base flow indices in both the calibration and validation periods.

5. Discussion

[32] When comparing calibrated parameter sets, there is general agreement among the automatic and manual calibrated parameter values, except for UZSN and INTFW. The difference in the INTFW values is the main reason for the difference in surface runoff predictions between the two calibration methods. INTFW determines the amount of water which enters the ground from surface detention storage and becomes interflow, as opposed to direct over-



(a)

(b)

Figure 8. Observed and simulated flow exceedance fraction over the calibration period using (a) automatic and (b) manual methods.

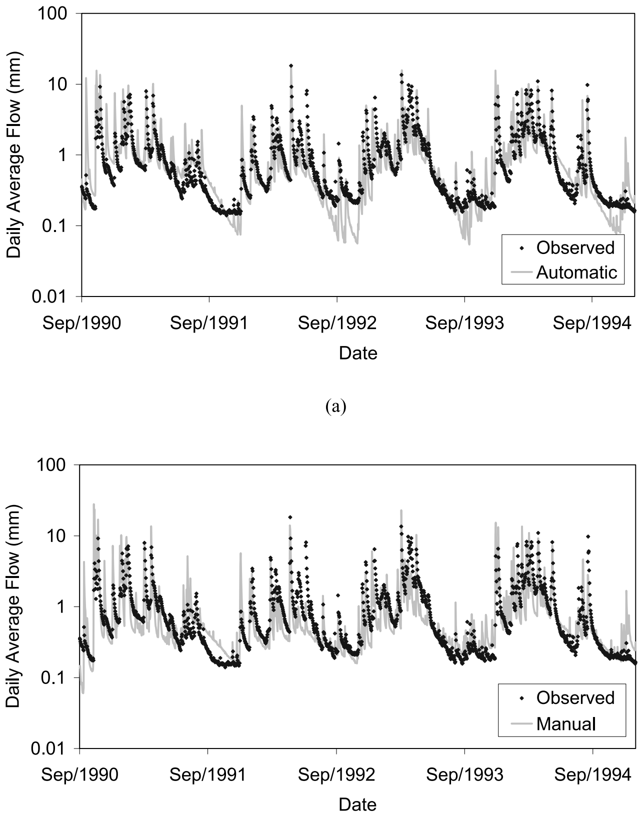
Table 7. Summary of Calibration Statistics for North River Watershed

Measures	Observed	Automatic Method		Manual Method		HSPEXP Model Performance Criteria, %
		Simulated	Percent Error	Simulated	Percent Error	
Total runoff, mm	1618.6	1571.2	−2.93	1463.3	−9.59	10
Total of highest 10% flows, mm	699.1	706.6	1.07	694.2	−0.71	15
Total of lowest 50% flows, mm	233.6	259.6	11.15[a]	254.3	8.86	10
Summer flow volume, mm	233.3	233.2	−0.07	230.1	−1.37	. . .
Winter flow volume, mm	362.6	386.1	6.46	370.3	2.12	. . .
Seasonal volume error, %	. . .	. . .	6.53	. . .	3.49	10
Average of storm peaks, m^3/s	68.4	55.1	−19.52[a]	59.1	−13.66	15
Summer storm volume, mm	37.2	39.9	7.31	41.1	10.73	15
		Automatic Method		Manual Method		
Coefficient of determination (R^2)		0.51		0.49		
Coefficient of efficiency (E)		0.35		0.29		
$RMSE$, mm		1.50		1.54		

[a] Violates the HSPEXP model performance criteria.

land flow and upper zone storage. Since the automatic method calibrated the INTFW value higher, the surface runoff was lower than that predicted with the manually calibrated parameters.

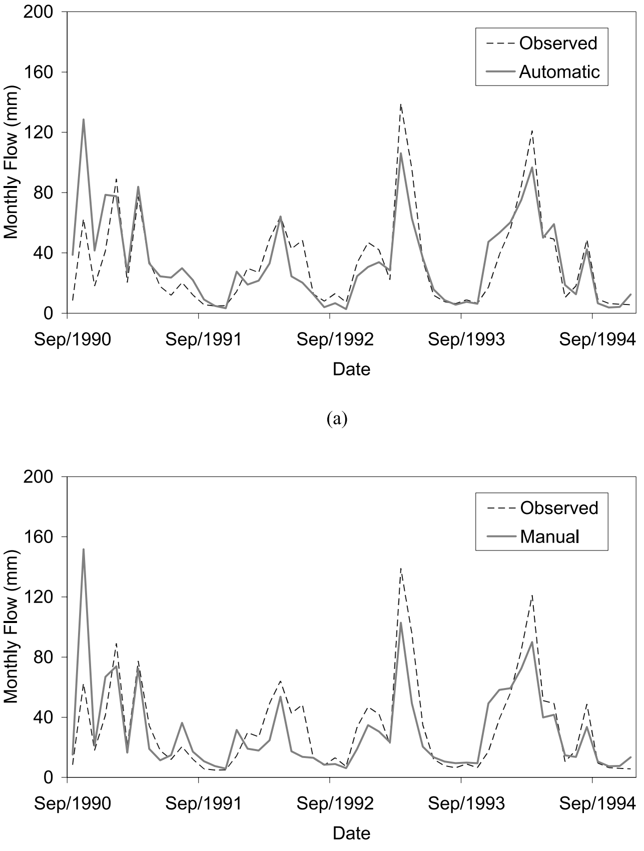
[33] The suitability of automatic calibration of HSPF using PEST as an alternative to manual calibration assisted by HSPEXP depends largely on the user's ultimate goal for the HSPF model. If compliance with HSPEXP model performance criteria is necessary, due either to regulation or some other client or project constraint, then the automatic method using PEST may not be sufficient to use on its own. The manual calibration assisted by HSPEXP satisfied all HSPEXP model performance criteria for both the calibration and validation periods. While PEST can use HSPEXP model performance criteria in the objective functions, it



(a)

(b)

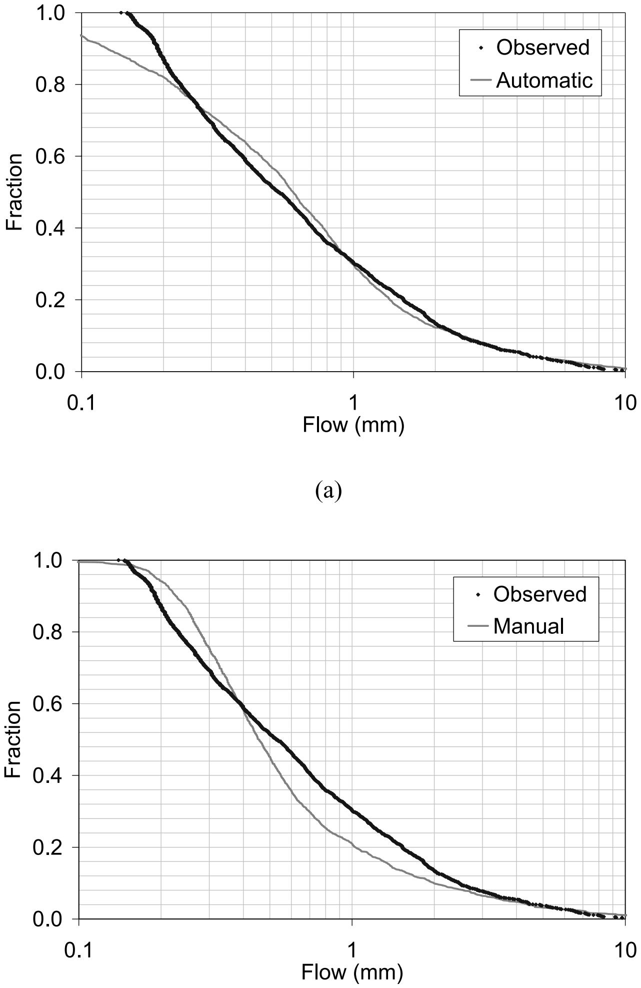
Figure 9. Observed and simulated daily flows over the validation period using (a) automatic and (b) manual methods.



(a)

(b)

Figure 10. Observed and simulated monthly flows over the validation period using (a) automatic and (b) manual methods.



(a)

(b)

Figure 11. Observed and simulated flow exceedance fraction over the validation period using (a) automatic and (b) manual methods.

cannot force the optimized results to satisfy those HSPEXP model performance criteria. In the study reported here, the automatic calibration parameter set violated two HSPEXP model performance criteria during calibration and one criterion during validation (Tables 7 and 8). If meeting HSPEXP model performance criteria is necessary, automatic calibration could still provide a good starting point for an expedited hydrology calibration, reducing the number of manual calibration runs needed to meet HSPEXP model performance criteria.

[34] If one is more concerned with conventional measures of goodness-of-fit as measures of model performance, it may be preferable to utilize PEST to calibrate HSPF. In this study, the automatic calibration parameter set produced simulation results with better conventional goodness-of-fit measures R^2, E, and $RMSE$. This result is not surprising since the manual calibration method assisted by HSPEXP is not intended to optimize these coefficients, whereas with the automatic method using PEST daily runoff was the basis of all the subobjective functions; daily runoff is also closely related to the goodness-of-fit statistics. Additionally, whereas the automatic calibration could end before all HSPEXP model performance criteria were satisfied, the manual calibration could not. Thus the additional steps needed to bring all HSPEXP model performance criteria into compliance during the manual calibration can cause some HSPEXP model performance variables to actually increase. For example, in the manual calibration of North River, a consequence of bringing the storm peaks error into compliance with the HSPEXP model performance criteria was an increase in the magnitude of the total volume error.

[35] Overall, both calibrated parameter sets over-predicted low flows, under-predicted middle-flows and under-predicted base flow indices during both the calibration and validation periods. The observed base flow ranged from 0.092 to 4.86 mm; the under-predicted middle-flows around 1 mm (Figures 8 and 11) resulted in the base flow indices being under-predicted.

[36] Some important factors that greatly affect surface water quality modeling, e.g., amount of surface runoff (Table 9), storm peaks (Table 7), and summer storm volumes (Table 8), were noticeably different between the automatic and manual calibration methods. The manual parameter set provided a better estimate of storm peaks and storm volumes than the automatic parameter set. Determining which method provides a better estimate of the breakdown between surface runoff and interflow is not straightforward (Table 9). The fact that a discrepancy exists means that modeled surface erosion and transport of pol-

Table 8. Summary of Validation Statistics for North River Watershed

Measures	Observed	Automatic Method		Manual Method		HSPEXP Model Performance Criteria, %
		Simulated	Percent Error	Simulated	Percent Error	
Total runoff, mm	1722.8	1720.3	−0.14	1597.2	−7.29	10
Total of highest 10% flows, mm	763.2	823.5	7.90	823.5	7.90	15
Total of lowest 50% flows, mm	226.0	230.6	2.03	240.5	6.42	10
Summer flow volume, mm	217.0	206.2	−4.96	198.1	−8.71	...
Winter flow volume, mm	516.4	530.9	2.79	517.1	0.14	...
Seasonal volume error, %	...	...	7.76	...	8.85	10
Average of storm peaks, m^3/s	50.4	57.7	14.50	57.3	13.62	15
Summer storm volume, mm	21.6	26.4	22.35[a]	21.3	−1.18	15

	Automatic Method	Manual Method
Coefficient of determination (R^2)	0.53	0.38
Coefficient of efficiency (E)	0.39	0.30
$RMSE$, mm	1.23	1.57

[a] Violates the HSPEXP model performance criteria.

Table 9. Flow Partitioning for the Calibration and Validation Periods

	Calibration			Validation		
Average Annual Flow	Observed	Automatic	Manual	Observed	Automatic	Manual
Total annual flow, mm	323.7	314.2	292.7	382.8	382.3	354.9
Surface flow, mm	...	6.8	30.9	...	9.9	49.6
Interflow, mm	...	154.6	124.2	...	205.6	165.6
Surface flow + Interflow, mm	146.7	161.4	155.1	171.5	215.5	215.2
Base flow, mm	176.9	152.9	137.5	211.3	166.7	139.7
Base flow index	0.55	0.49	0.47	0.55	0.44	0.39

lutant loads from the soil surface to the stream will be different for the two calibration methods.

6. Summary and Conclusions

[37] The objective of this research was to compare the calibrated parameter sets for HSPF resulting from an automated calibration with PEST and a manually assisted calibration with HSPEXP. The hydrologic calibration of HSPF was performed for the North River watershed in Virginia (United States) using the automatic and manual calibration methods. The hypothesis of this study was that automatic calibration using PEST is a viable alternative to manual calibration assisted by HSPEXP.

[38] PEST is a model-independent parameter optimization program that minimizes an “objective function” quantifying the misfit between model outputs and corresponding observed data. PEST has been widely used in the field of groundwater modeling, but there have been very few applications of PEST in the field of surface water modeling. HSPEXP is currently the most widely used software in the calibration of the hydrologic component of HSPF. HSPEXP is a decision support system that guides the user through the calibration process by offering advice based on expert users’ experience.

[39] For this study, multiple-objective functions were developed for use in PEST based on the HSPEXP model performance criteria. The development of these multiobjective functions represents an advancement over past applications of PEST to surface water modeling. Both the automatic and manual methods used observed data from a 5-year period (September 1985 to August 1990) for calibration and then an independent 4.5-year period (September 1990 to December 1994) for validation. The calibrated and validated results obtained from both calibration methods were compared using HSPEXP model performance criteria, R^2, E, $RMSE$ and base flow indices.

[40] The automatic calibrated parameter set produced by PEST satisfied most of the HSPEXP model performance criteria and performed better than the manual calibration with respect to R^2, E, and $RMSE$ for both the calibration and validation periods. The results of this research suggest that PEST may be a suitable alternative to HSPEXP for calibration of the hydrologic component of HSPF. However, further research is needed to investigate the possible ramifications of the differences between the HSPEXP and PEST calibrated hydrology parameter sets on water quality modeling.

[41] One of the main advantages of PEST is that the objective function can be tailored to one’s modeling objectives. In this study, for the comparison with manual calibration using HSPEXP, the objective functions were chosen according to the HSPEXP model performance criteria. However, if low flows are important, objective functions could be developed that focus on low flows rather than other types of flow regimes. As another example, if erosion and sediment transport are to be modeled, total flow volumes may be less important and spring and summer storms become more important. Once again, the objective functions used in PEST could be altered to focus on the spring and summer storms. The same would be true if the calibrated hydrologic parameter set was to be used to simulate water quality constituents. Another advantage of automatic calibration over manual calibration is that PEST could be used directly to calibrate erosion/sediment and water quality parameter sets. A suite of objective functions would need to be developed for these applications. This would lead to methods for comparing the overall quality of calibrations across watersheds in a more objective manner. However, further research is necessary to develop the objective functions and the associated weights, since there is currently no guidance available for the use of PEST in these broader applications in surface water modeling. Considering the positive performance reported in this research and the many possible applications of this automatic calibration method, PEST offers a new frontier for improving the field of surface water modeling.

References

Benham, B. L., K. M. Brannan, S. M. Morris, G. R. Yagow, and R. W. Zeckoski (2005), Bacteria total maximum daily load development for North River, Va. Dep. of Environ. Qual., Richmond. (Available at http://www.deq.virginia.gov/tmdl)

Bicknell, B. R., J. C. Imhoff, J. L. Kittle Jr., T. H. Jobes, and A. S. Donigan Jr. (2001), Hydrological Simulation Program–Fortran (HSPF): User’s manual for release 12, U.S. Environ. Prot. Agency, Athens, Ga.

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.

Brazil, L. E. (1988), Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. dissertation, 217 pp., Colo. State Univ., Fort Collins.

Doherty, J. (2004), PEST: Model-independent parameter estimation, User manual: 5th ed., Watermark Numer. Comput., Brisbane, Queensl., Australia.

Doherty, J., and J. M. Johnston (2003), Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, *39*, 251–265.

Environmental Systems Research Institute (2003), ArcGIS 8.3, Redlands, Calif.

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.

Hummel, P., J. Kittle Jr., and M. Gray (2005), WDMUtil version 2.0 user’s manual, Off. of Water, U.S. Environ. Prot. Agency, Washington, D. C.

(Available at http://www.epa.gov/waterscience/basins/b3docs/wdmutil.zip)

Lawson, L. G. (2003), Guidance memo 03-2012: HSPF model calibration and verification for bacteria TMDLs, Va. Dep. of Environ. Qual., Richmond. (Available at http://www.deq.virginia.gov/waterguidance/pdf/032012.pdf)

Lumb, A. M., R. B. McCammon, and J. L. Kittle (1994), Users manual for an EXPert System (HSPEXP) for calibration on the Hydrological Simulation Program–Fortran, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *94-4168*.

Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, *235*, 276–288.

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.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, *11*(2), 431–441.

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

National Resource Conservation Service (2004), Regional hydraulic geometry curves, Natl. Water Manage. Cent., Little Rock, Ark. (Available at http:/wmc.ar.nrcs.usda.gov/technical/HHSWR/Geomorphic/index.html)

Sloto, R. A., and M. Y. Crouse (1996), HYSEP: A computer program for streamflow hydrograph separation and analysis, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *96-4040*, 46 pp.

Southeast Regional Climate Center (2004), Dale Enterprise, Virginia (442208): Period of record monthly climate summary—1 August 1948 to 31 March 2004, S.C. Dep. of Nat. Resour., Columbia. (Available at http://cirrus.dnr.state.sc.us/cgi-bin/sercc/cliMAIN.pl?va2208)

Staley, N., T. Bright, R. W. Zeckoski, B. L. Benham, and K. M. Brannan (2006), Comparison of HSPF outputs using FTABLES generated with field survey and digital data, *J. Am. Water Resour. Assoc.*, *42*(5), 1153–1162.

U.S. Environmental Protection Agency (1999), BASINS technical note 5: Using HSPEXP with BASINS/NPSM, *EPA-823-R-99-010*, Off. of Water, Washington, D. C. (Available at http://www.epa.gov/waterscience/basins/tecnote5.pdf)

U.S. Environmental Protection Agency (2000), BASINS technical note 6: Estimating hydrology and hydraulic parameters for HSPF, *EPA-823-R-00-012*, Off. of Water, Washington, D. C.

U.S. Environmental Protection Agency (2001), The national costs of the total maximum daily load program, *EPA 841-D-01-003*, Washington, D. C. (Available at http://www.epa.gov/owow/tmdl/coststudy/coststudy.pdf)

U.S. Geological Survey (2005), Daily streamflow for the nation, Washington, D. C. (Available at http://nwis.waterdata.usgs.gov/usa/nwis/discharge)

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

B. L. Benham, K. M. Brannan, S. M. Kim, and R. W. Zeckoski, Biological Systems Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060-0303, USA. (kbrannan@vt.edu)

J. Doherty, Watermark Numerical Computing, 336 Cliveden Avenue, Corinda 4075, Queensland, Australia.