Assessing the Performance of HSPF When Using the High Water Table Subroutine to Simulate Hydrology in a Low-Gradient Watershed

M. Scott Forrester[1], Brian L. Benham[*2], Karen S. Kline[2], Kevin J. McGuire[3]

[1]Hatch Mott MacDonald Millburn, NJ, 07041, USA

[2]Dept. Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24060, USA

[3]Dept. Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA 24060, USA

[*2]benham@vt.edu

***Abstract*- Modeling groundwater hydrology is critical in low-gradient, high water table watersheds where ground-water is the dominant contribution to streamflow. The Hydrological Simulation Program-FORTRAN (HSPF) model has two different subroutines available to simulate ground water, the traditional groundwater (TGW) subroutine and the high water table (HWT) subroutine. The HWT subroutine has more parameters and requires more data but was created to enhance model performance in low-gradient, high water table watershed applications. The objective of this study was to compare the performance and uncertainty of the TGW and HWT subroutines when applying HSPF to a low-gradient watershed in the Coastal Plain of northeast North Carolina. Monte Carlo simulations were performed to generate data needed for model performance comparison. Both models performed well when simulating the 10% highest daily average flow rates. However, neither model performed well when simulating the 50% lowest daily average flow rates. The HWT model significantly outperformed the TGW model when simulating daily average flow over the full three-year simulation period, an indication that the HWT model out-performed the TGW over the full range of simulated flows. Model uncertainty was assessed using the Average Relative Interval Length (ARIL) metric. The HWT model exhibited slightly more combined model structure and parameter uncertainty than the TGW model. Based on the results, the HWT subroutine is preferable when applying HSPF to a low-gradient watershed and the accuracy of simulated stream discharge is the paramount concern.**

Keywords- Modeling; HSPF; High Water Table; Uncertainty

I. INTRODUCTION

Coupled hydrological and water quality models have evolved over time increasing in capability and complexity [1, 2, 3]. The Hydrological Simulation Program-FORTRAN (HSPF) is an example of a complex watershed-scale, process-based, lumped-parameter model that combines physical data and a substantial set of parameters to simulate hydrology and pollutant fate and transport [4, 5, 6]. HSPF simulates the movement of water, sediment, and a wide range of water quality constituents (e.g., nutrients, sediment, bacteria) on pervious and impervious surfaces, in the soil, and in streams and well-mixed reservoirs [7, 8], and is frequently used when conducting Total Maximum Daily Load (TMDL) studies in the U.S. [9, 10]. A TMDL is a quantitative representation of the contributions of a particular pollutant to a water body that specifies the pollutant reductions needed to restore water quality [11]. HSPF has been used to develop TMDLs for pollutants ranging from fecal indicator bacteria to PCBs [10, 11, 12]. While effectively characterizing all hydrologic processes is critical when developing a watershed–scale model, accurately characterizing groundwater contributions to in-stream flows can be challenging.

While groundwater contributions to in-stream flow can be significant in any topography, it is often the dominant component of streamflow in areas with relatively flat topography and accompanying high water tables. Effectively characterizing the groundwater component of the hydrologic budget is critical to developing an accurate model, especially when the model is applied to water quality improvement efforts like TMDLs [13, 14]. Reference [15] found that adding a specific groundwater routine to their modelling schema improved hydrology simulations in a watershed that exhibited a high groundwater table. The HSPF model provides two alternative groundwater subroutines that the user may choose from, the traditional groundwater subroutine (TGW) and the high water table (HWT) subroutine. The HWT subroutine was developed for use in areas dominated by wetlands and/or low-gradient watersheds [7, 15].

The TGW subroutine is used most frequently in HSPF applications. When using this subroutine, the groundwater contribution to in-stream flow is calculated as fraction of the active groundwater storage volume [16]. The active groundwater storage (AGWS) parameter represents a shallow groundwater aquifer that readily contributes to both evapotranspiration and baseflow [17]. Calibrated parameters determine how much of the groundwater in the AGWS is lost to evapotranspiration (AGWETP), contributes to baseflow (AGWRC), or percolates into deep aquifers (DEEPFR) [7]. In the TGW subroutine, water that percolates below the AGWS zone, into the deep aquifer (DEEPFR), is lost to a separate, inactive storage (IGWI) and does not contribute to baseflow or evapotranspiration [7].

The HWT subroutine involves a more sophisticated conceptualization of the subsurface profile and a more complex handling of groundwater, Fig 1. The HWT subroutine calculates a groundwater elevation variable (GWEL) at each time step to determine how groundwater interacts within various groundwater storage volumes and the overall hydrologic budget calculation. The HWT subroutine uses two, what are termed, influence elevations (UELV and LELV) to separate the sub-surface groundwater storage into zones or regions, as they are called in the HSPF documentation – an upper zone (Region 3), a middle zone (Region 2), and a lower zone (Region 1) [7]. Region 1 is assumed to be saturated and water in Region 1 does not interact with the unsaturated zone (above GWEL). Groundwater in Region 1 is quantified by the AGWS parameter. When GWEL falls within Region 1 the HWT subroutine operates no differently than the TGW subroutine. Region 2 lies above (more shallow) than Region 1 and is defined by the Lower Zone Influence Elevation (LELV) parameter. If the GWEL rises into Region 2, groundwater interacts with the lower zone storage (LZS). Water that enters LZS is assumed to be held in place by cohesive forces, meaning it cannot percolate back into the groundwater, it can, however, contribute to evapotranspiration (AGWETP) or baseflow (AGWRC). When the GWEL rises into Region 3 (most shallow), above the Upper Zone Influence Elevation (UEL), groundwater can contribute to evapotranspiration (AGWETP) or as interflow (IFWS). If the water table (i.e., GWEL) rises to the surface, this water is added to surface storage (SURS) and can potentially contribute to runoff [7]. In the HWT subroutine, subsurface storage volumes are characterized by a number of porosity parameters; PCW represents the micropore and is constant for all three regions, PGW represents the porosity present in macropores in Regions 1 and 2, and UPGW represents porosity of the macropores in Region 3. In practice, the porosities for the three regions (PGW, PCW, UPGW) are typically quantified using available soil profile data and averaged over the different land-uses and subwatersheds [7].

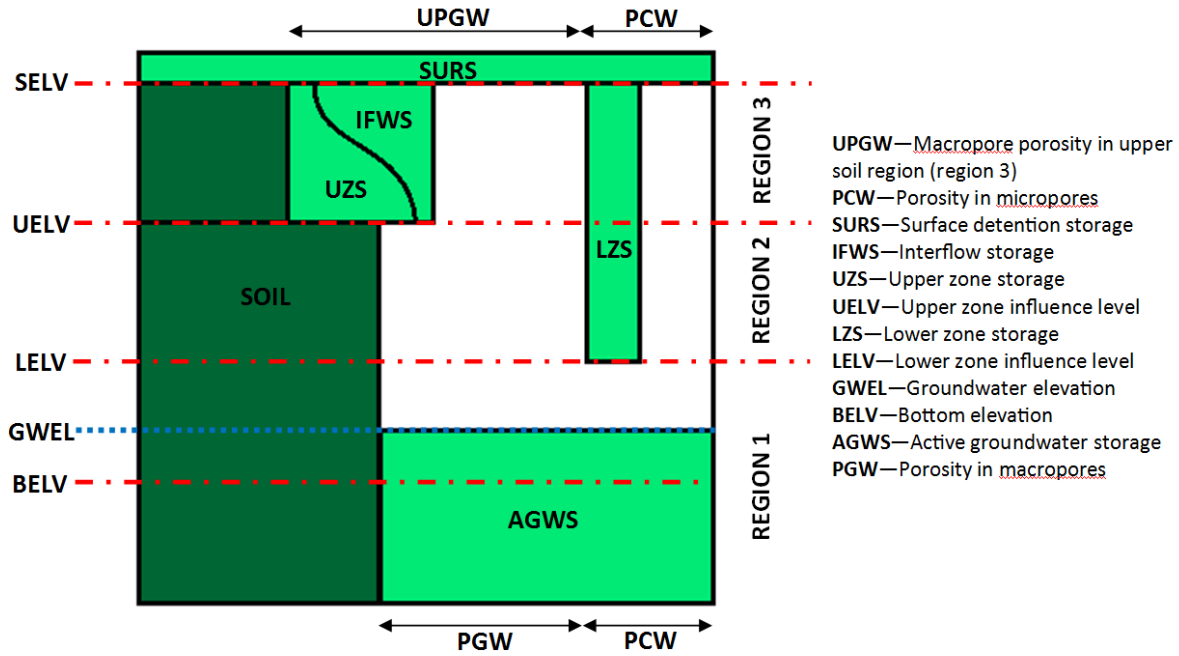

Fig. 1 Schematic of the Soil Profile Representation Used in the HWT Subroutine (Bicknell *et al.*, 2001)

While the HWT subroutine was developed specifically for use in low-gradient watersheds that may exhibit high water table behaviour, there is currently no literature comparing the performance of these two TGW and HWT groundwater subroutines available in HSPF. The objective of this research was to compare the performance, uncertainty, and utility of use of the TGW and HWT subroutines when using HSPF to simulate stream discharge in a low-gradient watershed in a Coastal Plain watershed in northeast North Carolina and to determine if either subroutine exhibited a demonstrable advantage over the other.

II. METHODS

A. Study Site

The Ahoskie Creek watershed (14,700 ha) in northeastern North Carolina was selected for this study (Fig. 2). Ahoskie Creek lies in the Lower Coastal Plain physiographic province, a region characterized by flat surface slopes (<2%) and channel slopes (<0.1%) [18]. Ahoskie Creek is in the Chowan River basin which drains to the Albemarle Sound. The dominant land uses are forest (50.7%) and cropland (22.9%) [19]. The stream reaches in the watershed were channelized between 1969 and 1972 to improve drainage. Prior to channelization, the watershed did not have clearly discernible channels resulting in poor drainage and frequent flooding events [20]. The low-gradient, documented flooding events, and presence of wetlands indicated a high, active water table making the Ahoskie Creek watershed a good candidate for this study. In addition, a long record of discharge was available from a USGS gauge on the Ahoskie.

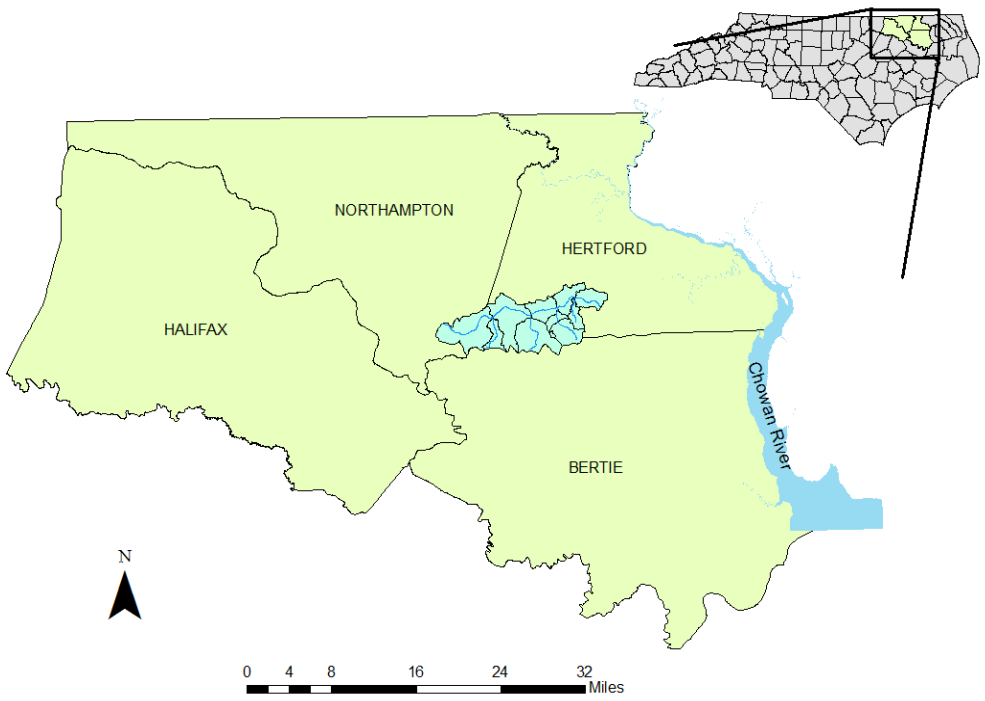

Fig. 2 Location of the Ahoskie Creek Watershed, North Carolina, USA

B. Model Development and Parameterization A

1) Watershed Characterization and Observed Data:

The HSPF model relies on a combination of remotely sensed data and field-based observations to characterize the watershed and parameterize the model. Using digital raster graphics (DRG) and digital elevation model (DEM) data [21, 22] the Ahoskie Creek watershed was manually delineated into ten (10) subwatersheds, with an average subwatershed area of 1559 ha (Fig. 3). The land uses for Ahoskie Creek were derived from the National Land Cover Database (NLCD) [23]. The NLCD land uses present in the watershed were aggregated into five broad land use categories: forest, 50.7%; cropland, 22.9%; pasture, 11.3%; residential, 2.2%; and wetlands, 12.9% (Fig 4). When using HSPF, each subwatershed can include pervious land segments (PLS) and/or impervious land segments (IPS). Watershed characterization involves parameterizing the various PLSs and IPSs for each subwatershed. The Soil Survey Geographic database (SSURGO) (USDA-NRCS) was used to generate the HSPF infiltration parameters (INFILT) used in both models and soil porosity parameters (PCW, PGW, UPGW) used only in the HWT subroutine. Unique values for INFILT, PCW, PGW, and UPGW were assigned to each PLS by taking an area weighted average of the soils intersecting the five different land uses.

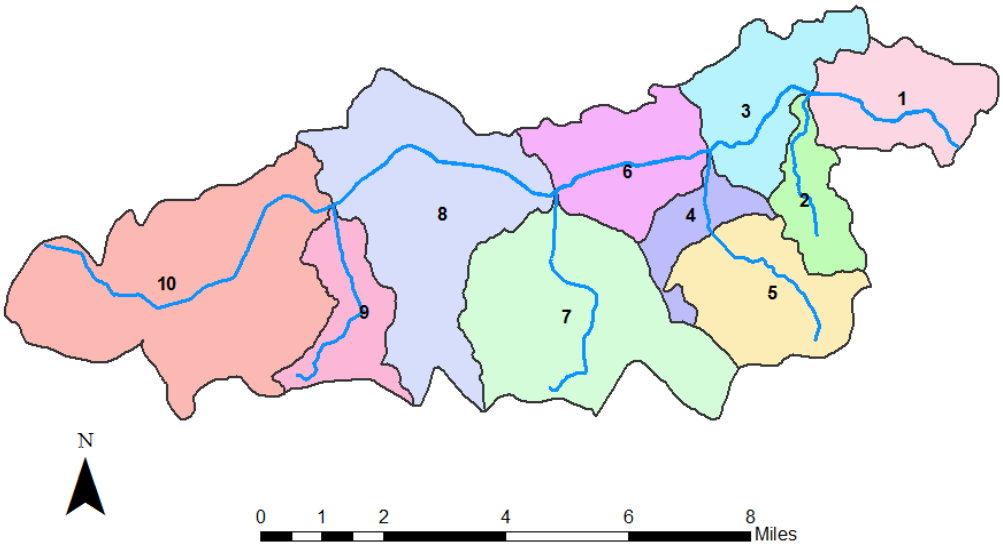

Fig. 3 Ahoskie Creek Subwatersheds

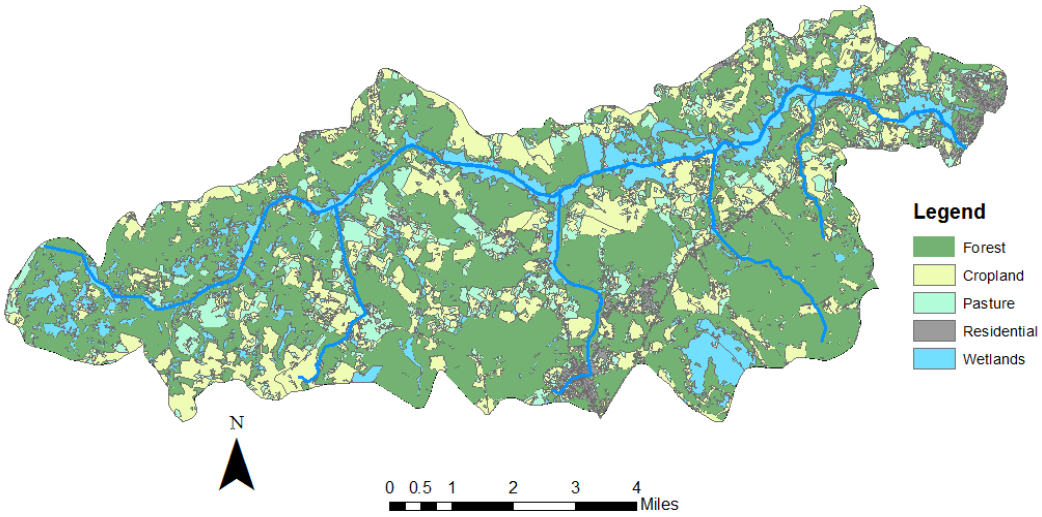

Fig. 4 Ahoskie Creek Watershed Land Use Distribution

HSPF uses stage-volume-discharge relationships called Hydraulic Function Tables (FTABLES) at the outlet of each subwatershed to quantify discharge. For this study, FTABLES were developed using surveyed stream cross-section data. The cross-sectional profiles were assumed to be prismatic within a subwatershed. The length of each reach and the cross-sectional profiles formed stage-volume relationships. Manning's equation was then used for each reach, taking into account the surface roughness, slope of the subwatershed, and cross-sectional profile, to calculate discharge and complete the stage-volume-discharge relationship. The Manning's roughness coefficient (n) was determined for each reach by examining the streambed and banks and comparing them to stream channel reference photos [24].

Observed daily flow data (1950 to 2011) were obtained from the USGS gaging station located on Ahoskie Creek in Ahoskie, NC (02053500). Precipitation and air temperature data were obtained primarily from the National Weather Service Cooperative Observer Program (COOP) station in Murfreesboro, NC (315996) that has an average annual rainfall of 116.05 cm. The Murfreesboro station is located 16 km northwest of the Ahoskie Creek watershed and is the closest active COOP station to the watershed. Gaps or inconsistencies in the Murfreesboro climate record were filled using data from Roanoke Rapids, NC (317319) and Wakefield, VA (448800) located 64 km northwest and 72 km north, respectively, from the Ahoskie Creek watershed. Wind speed, percent sun, and dew point temperature were all obtained from the Norfolk International Airport COOP station (446139).

2) Monte Carlo Simulations:

To generate the data for the model performance comparison and uncertainty analysis, one hundred thousand (100,000) Monte Carlo (MC) simulation runs were performed for each model variation (i.e., TGW and HWT). To perform the MC simulations and isolate the performance of the groundwater subroutines, the parameters used to characterize each groundwater subroutine were treated as stochastic, while the parameters common to both models, but external to the groundwater subroutines, were treated as deterministic.

Deterministic parameters values were initially set using the HSPF guidance document, BASINS Technical Note 6 [22]. Those initial parameter values were refined through model calibration using only the TGW subroutine model. Dummy variables were used during the calibration process for the TGW parameters that would ultimately be treated as stochastic for the MC runs. Those dummy parameter values were based on an HSPF model that had been previously calibrated for a nearby watershed. The calibration period was from 1 January 1990 through 31 December 2009. Model calibration sufficiency was assessed by comparing the simulated and observed stream discharge at an hourly time step. Calibration was deemed satisfactory when the Ahoskie Creek TGW model output met the widely used HSPF calibration assessment criteria shown in Table I [26]. The values for the calibrated deterministic model parameters are shown in Table II. Model validation was not performed given that the focus of this research was to compare the two HSPF groundwater subroutines performance against one another.

TABLE I HSPF MODEL ASSESSMENT CALIBRATION CRITERIA AND THE AHOSKIE CREEK CALIBRATED MODEL ERROR

Flow Statistic	Calibration Criteria Error (%)	Ahoskie Creek Calibrated Model Error (%)
Error in total runoff (mm)	± 10	9.5
Error in low flow recession (%)	± 0.01	0.01
Error in 50% lowest flows (mm)	± 10	9.2
Error in 10% highest flows (mm)	± 15	7.2
Error in storm peak (m^3/s)	± 15	-12.1
Seasonal volume error (%)	± 10	10.0
Summer storm volume (mm)	± 15	13.7

TABLE II HSPF MODEL ASSESSMENT CALIBRATION CRITERIA AND THE AHOSKIE CREEK CALIBRATED MODEL ERROR

Parameter	Description, Units	Calibrated Value
Pervious Land Uses (PERLND)		
INFILT	Index to infiltration capacity, in/hr	0.023-0.109[†]
LSUR	Length of overland flow, in/hr	452-513[†]
SLSUR	Slope of overland flowplane	0.0039-0.0322[a]
KVARY	Groundwater recession variable, 1/in	0
PETMAX	Temp below which ET is reduced, deg. F	40
PETMIN	Temp below which ET is set to zero, deg. F	35
INFEXP	Exponent in infiltration equation	2
INFILD	Ratio of max/mean infiltration capacities	2
CEPSC	Interception storage capacity, inches	0.06-0.27[‡]
NSUR	Manning's n (roughness)	0.45 forest; 0.30 pasture and residential; 0.27 crop; 0.40 wetlands
Impervious Land Uses (IMPLND)		
LSUR	Length of overland flow, feet	497
SLSUR	Slope of overland flowplane	0.0116
NSUR	Mannings' n (roughness)	0.1
RETSC	Retention/interception storage capacity, inches	0.1
PETMAX	Temp below which ET is reduced, deg. F	40
PETMIN	Temp below which ET is set to zero, deg. F	35
Stream Reaches and Reservoirs (RCHRES)		
KS	Weighting factor for hydraulic routing	0.5

[†]Varies with land use, [‡]Varies by month and with land use

The MC simulations that were used to generate data for the model comparison were executed using a Visual Basic.NET program (VBP) script written specifically for this study. The VBP script populated the groundwater subroutine parameter values in the fixed-format HSPF User Control Input (UCI) file by sampling uniform parameter distributions, whose ranges were based on the "typical" parameter values provided in BASINS Technical Note 6 [25]. The HWT subroutine used the same stochastic parameters used in the TGW model plus three additional parameters – the surface runoff recession coefficient (SRRC), the surface runoff exponent (SREXP), and the interflow storage capacity (IFWSC). The SRRC parameter adjusts surface runoff and is a function of surface storage. The SREXP is a multiplicative factor used in the calculation of surface runoff. And IFWSC is the maximum interflow storage capacity when the ground-water elevation (GWEL) rises above the upper influence elevation (UELV). These HWT-specific parameters were assigned uniform distributions with initial distribution ranges based on HSPF user guidance [7]. The maximum value for the IFWSC parameter was set at of 6.8 to ensure model stability. The upper bound of the SREXP was set to 2.0 to prevent excess water from being attributed to surface runoff. Table III shows the TGW and HWT subroutine parameters that were treated as stochastic for this study and uniform distribution limits.

TABLE III STOCHASTIC PARAMETERS AND RANGES FOR THE TRADITIONAL GROUND-WATER (TGW) SUBROUTINE

Parameter	Parameter Description, Units	Minimum	Maximum
TGW and HWT Subroutines			
INTFW[†]	Interflow inflow	1.000	3.000
IRC[†]	Interflow recession coefficient, 1/day	0.100	1.000
UZSN[†]	Upper zone nominal storage, in.	3.000	8.000
LZSN	Lower zone nominal storage, in.	0.500	0.700
AGWRC	Ground-water recession rate, 1/day	0.920	0.990
DEEPFR[‡]	Fraction of ground water lost to deep, inactive aquifers	0.001	0.050
BASETP[‡]	Fraction of remaining ET potential satisfied by baseflow	0.001	0.050
AGWETP	Fraction of remaining ET potential satisfied by active ground-water storage	0.001	0.200
LZETP - Forest	Lower zone ET potential	0.600	0.800
LZETP - Cropland	Lower zone ET potential	0.500	0.700
LZETP - Pasture	Lower zone ET potential	0.400	0.600
LZETP - Residential	Lower zone ET potential	0.200	0.500
LZETP - Wetlands	Lower zone ET potential	0.600	0.900
HWT Subroutine only			
SRRC	Surface runoff recession coefficient	0.010	0.990
SREXP	Surface runoff exponent	0.100	2.000
IFWSC	Max Interflow storage capacity	0.100	6.800

[†] Each land use was represented in HSPF with a separate parameter but all were given the same initial distribution.

[‡] The forest, cropland, pasture, and residential land uses were all represented in HSPF with one parameter. Wetlands were represented with a separate parameter but given same initial distribution.

The HSPF infiltration parameter (INFILT) has been shown to be a sensitive parameter [27, 28] and is frequently adjusted when calibrating HSPF. For this study, after the initial calibration, INFILT was held constant across all land uses. Holding INFILT constant further isolated the groundwater subroutines from the rest of the HSPF model. In the same vein, the soil porosity parameters (PCW, PGW, UPGW) used in the HWT model were held constant throughout the MC simulations. Again, this allows for more consistency between runs and isolation of the groundwater subroutines. It is important to note that while the baseline TGW model used to calibrate the deterministic parameters used lower zone evapotranspiration (LZETP) values that varied by both landuse and month, the two models used for subroutine comparison treated LZETP as a stochastic parameter and allowed LZETP to vary only by land use. Preventing the land use-specific LZETP values from varying by month reduced one source of variability between subroutine parameterization permitting a more direct subroutine performance and uncertainty comparison.

C. Model Performance Comparison

For the comparisons made in this study, the model was run on an hourly time step. A three-year model assessment period (1 January 2006 to 31 December 2009) was used to compare model performance with a one-year startup period (1 January 2005 to 31 December 2005) to allow the model's state variables to equilibrate. The response variable was average daily flow rate in cubic meters per second (m^3/s). Model performance was compared using three (3) metrics and the model response variable average daily flow rate (m^3s^{-1}). The first metric, the Nash-Sutcliffe efficiency (NSE), is a good measure of overall model performance over the three year simulation period and was calculated using Equation (1). NSE values can range from -∞ to 1 [29]. The NSE index was chosen because it is commonly used in hydrologic literature and incorporates both variation about the mean and total variance not described by the model [30]. A NSE value was calculated for each run of the two models.

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_o(t) - Q_m(t)\right)^2}{\sum_{t=1}^{T}\left(Q_o(t) - \overline{Q}_o\right)^2} \quad (1)$$

where

Q_o = observed daily average flow rate (m^3/s)

Q_m = simulated daily average flow rate (m^3/s)

t = time, in days

The second metric used to compare model performance provided a measure of how effectively the models simulated the lowest 50% of observed flows. The third metric used was a measure of how effectively the models simulated the highest 10% of observed flows. These metrics are often used by HSPF modelers who use the HSPEXP calibration assessment tool [31] to assess calibration adequacy. When computing the 50% lowest flows metric, HSPEXP sums the volume of the lowest 50% of observed flows and then sums the simulated flow volumes for those time intervals. The volume of the highest 10% flows is summed similarly. The percent error for both metrics was computed using Equation (2). Based on the calibration criteria (Table I), an error of +/- 10% is considered acceptable for the 50% low flow metric and +/- 15% is acceptable for the high flow metric. While the response variable considered in this study was daily average flow rate instead of flow volume, the HSPEXP guidance for both metrics was used to assess model performance. The performance metric results from the two models were compared using the non-parametric Wilcoxon Rank-Sum test.

$$\%Error = \frac{Simulated\ sum - Observed\ sum}{Observed\ sum} \quad (2)$$

1) Behavioural Parameter Set Selection:

The concept of behavioural versus non-behavioural parameter sets was first introduced by [32] and is used in the Generalized Likelihood Uncertainty Estimation (GLUE) method of estimation uncertainty [33]. Parameter sets resulting in model runs that fall above a threshold criteria are considered to be behavioural and provide an acceptable characterization of the watershed [33, 34]. The criteria for selecting the behavioural parameter sets for each model was determined by examining the cumulative distribution function (CDF) curves of each of the three previously discussed metrics. An upper inflection point in the curve indicated a threshold value below which the CDF value dropped sharply. Parameter sets for the model runs that yielded a metric value above the inflection point threshold value were considered behavioural and used in the model performance and uncertainty analyses. All other parameter sets were considered non-behavioural and were not considered further. This process of selecting behavioural parameter sets resulted in three different groups of behavioural parameter sets for both models.

2) Quantifying Model Uncertainty:

Uncertainty in hydrological modeling is unavoidable [34, 35] and comes from multiple sources (e.g., model input data,

model structure, and the variability of individual parameters [35, 36, 37, 38, 39]). Any robust comparison of model performance should attempt to quantify and compare relevant sources of uncertainty. The Average Relative Interval Length (ARIL) metric, Equation (3), introduced by [40], was used to quantify the uncertainty present in the two models resulting from groundwater subroutine model structure and parameter uncertainty. The ARIL metric provides a single measure for comparing the relative width of the confidence intervals throughout a specified time period. The 95% confidence intervals were formed using the 2.5% and 97.5% quantiles of the distribution of simulated flow values at each time step. The ARIL metric provides a single value that can be used to compare model performance. Smaller ARIL values indicate narrower confidence bands and less model uncertainty.

$$ARIL = \frac{1}{n}\sum \frac{Limit_{upper,t} - Limit_{lower,t}}{Q_{obs,t}} \tag{3}$$

where

n = number of time steps

$Limit_{upper,t}$ = 97.5% quantile for time step t

$Limit_{lower,t}$ = 2.5% quantile for time step t

$Q_{obs,t}$ = observed discharge value for time step t

3) Assessing Parameter Identifiability:

In conjunction with estimates of model uncertainty due to parameter variability, a regional sensitivity analysis of model parameters can be useful in determining parameters to which the model is more sensitive [34, 41, 42]. The most sensitive parameters will have an identifiable value or range that tends to yield better model results with respect to a particular model performance metric. Models with more identifiable parameters may be easier to calibrate to observed data [43, 44]. Parameter identifiability, especially in an over-parameterized model like HSPF, can be challenging during calibration [30]. Parameters that exhibited a trend towards a particular value or range were more sensitive within the model structure were considered to be "identifiable" [34, 45]. Parameter identifiability in this study was determined by comparing the predefined or *a priori*, uniform distributions of each parameter to their respective posterior distributions. Posterior parameter distributions were developed using the behavioural parameter sets discussed previously.

To assess parameter identifiability, the *a priori* and posterior parameter ranges were normalized from 0 to 1, and curvature of the normalized CDF posterior distribution was then compared to the normalized CDF *a priori*, uniform distribution. The CDF curve of a uniform distribution exhibits a 1 to 1 slope over the range from 0 to 1. The vertical distance between the two distributions was calculated at 1000 points along the normalized parameter range. Any parameter that showed a maximum vertical difference between the *a priori* and posterior CDF > 0.1, or 10% of the cumulative probability was considered to be identifiable (i.e. non-uniform). The selected parameter identifiability threshold is arbitrary but it allows for relative comparison of the number of identifiable parameters between the two models. Fig. 5 shows an example of a non-identifiable and an identifiable parameter. While other studies [43, 46] have used the Kolmogorov-Smirnov (KS) statistic to detect *a priori* and posterior parameter distribution differences, we chose to use the identifiability metric described above because the large parameter value sample size used in this study (10,000) yielded KS statistic p-values less than 0.05 for almost every parameter, even those whose posterior distributions appeared uniform when examined visually.

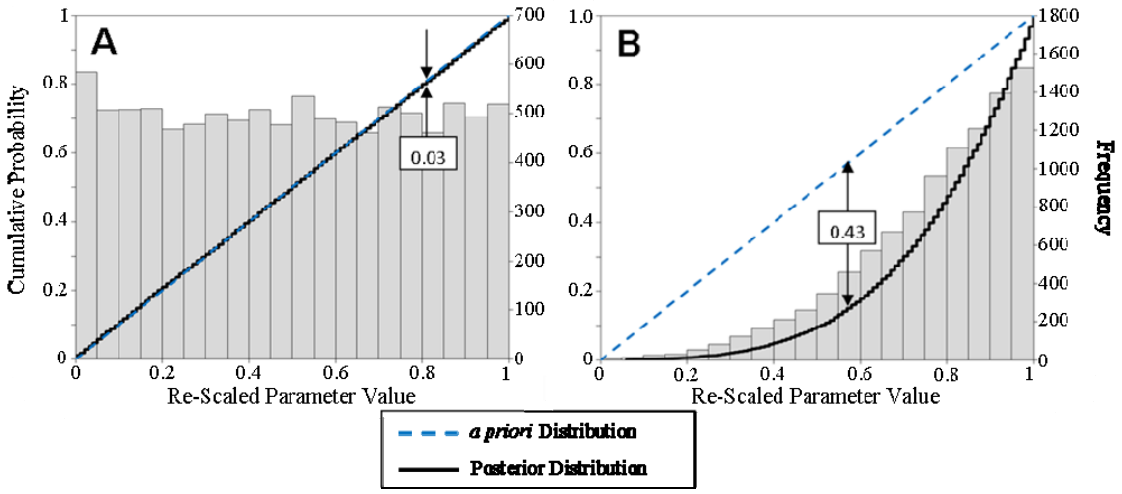

Fig. 5 Examples of a Non-Identifiable (A) and an Identifiable Parameter (B)

III. RESULTS AND DISCUSSION

A. Behavioural Parameter Sets

The model performance histograms and resulting CDF curves used to determine behavioural parameter set threshold based on each of the three model performance metrics are shown in Fig. 6. Four of the six CDF curves exhibited an upper inflection point where the sharp trend towards better model performance decreased at, or near, the 90th percentile. For consistency, the 90th percentile was chosen as the threshold value used to determine behavioural parameter sets for all three of the model performance metrics. Only model runs performing at or above the 90th percentile value for each of the metrics were used in the model performance and uncertainty analyses. Given 100,000 model runs, the 90th percentile threshold generated 10,000 behavioural parameter sets for each of the three model performance metrics.

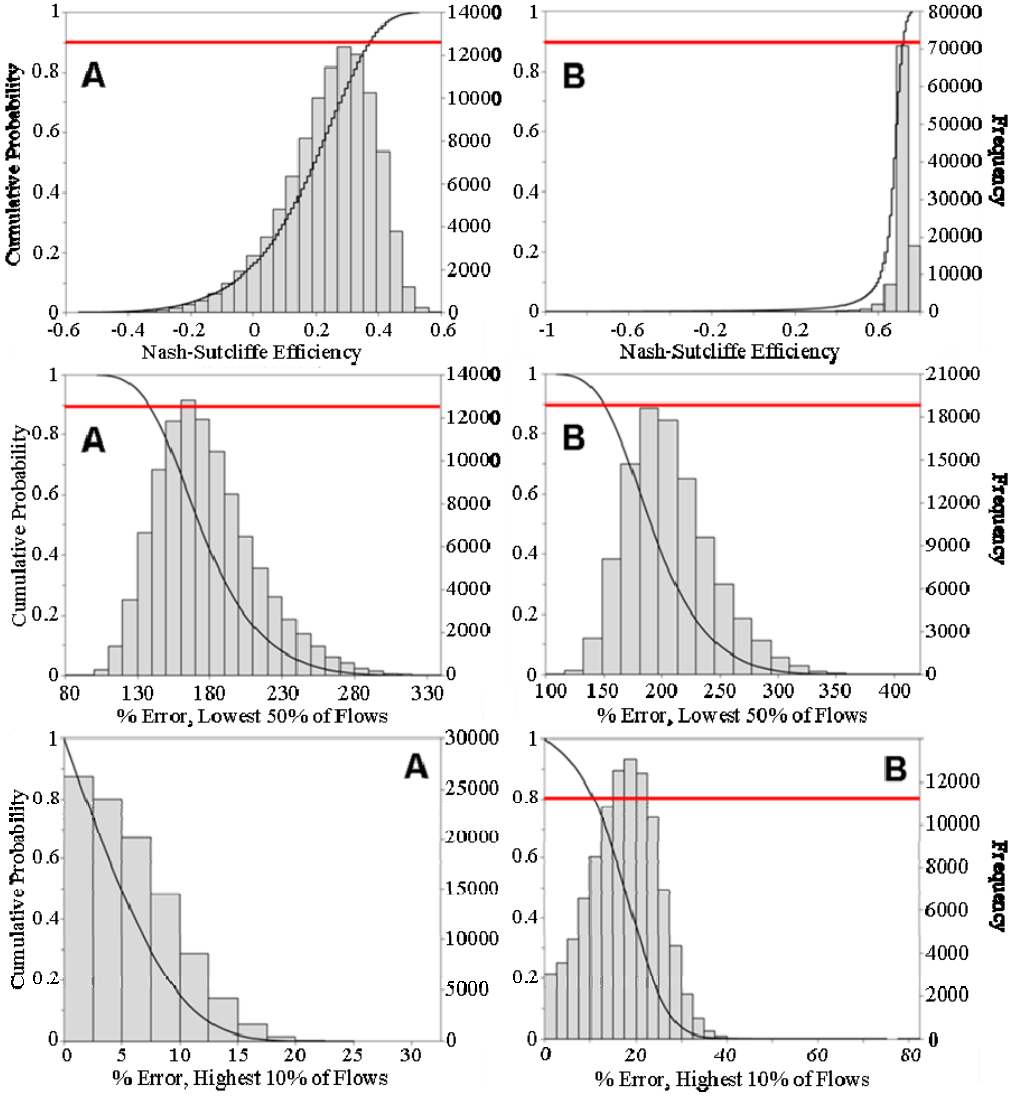

Fig. 6 CDF Curves and Histograms Used to Determine Behavioral Parameter Sets for the TGW (A) and HWT (B) Models. The solid, horizontal line indicates the inflection point

B. Model Performance Comparison

Using only the behavioural parameter sets, the performance of the two models was compared using the NSE (average daily flow rate) and the percent errors for the 50% lowest flows and 10% highest flows. As shown in Fig. 7, the HWT model significantly outperformed the TGW model with regards to the NSE metric. In fact, the maximum NSE achieved by the TGW model (0.60) was much less than the worst performing HWT behavioural model run (0.72). With respect to the 50% lowest average daily flow rate metric, the TGW model statistically outperformed the HWT model though both models did a poor job of simulating low flows; the percent errors were greater than 100% for both models. It is important to note, however, that many of the observed average daily flow rates used to assess this metric were quite low (<0.15 m^3/s), so even a small error in the simulated flow rate will result in a large percent error. With respect to the 10% highest flow metric, the TGW again statistically outperformed the HWT model. However, means of the 10% high flow distributions for both models, fell well within the

HSPEXP calibration criteria of +/-15%. Because the sample size for these comparisons was so large (10,000 values each), any difference in the means of the performance metric distributions for the two models was statistically significant when compared using the Wilcoxon Rank-Sum test, i.e. all p-values were less than 0.05. Therefore, from a practical perspective, both the TGW and HWT models performed equally poor when simulating low flows (mean percent errors of 129.0 and 138.1%, respectively). Further, they both performed equally well when simulating the 10% highest flows (mean percent errors of 0.5 and 3.8%, respectively). But, when model performance is compared over the entire observed flow range for the assessment period using the NSE metric, the HWT model clearly outperformed the TGW model when predicting daily average flow rate. The mean HWT NSE value was some 70% greater than the TGW NSE value.

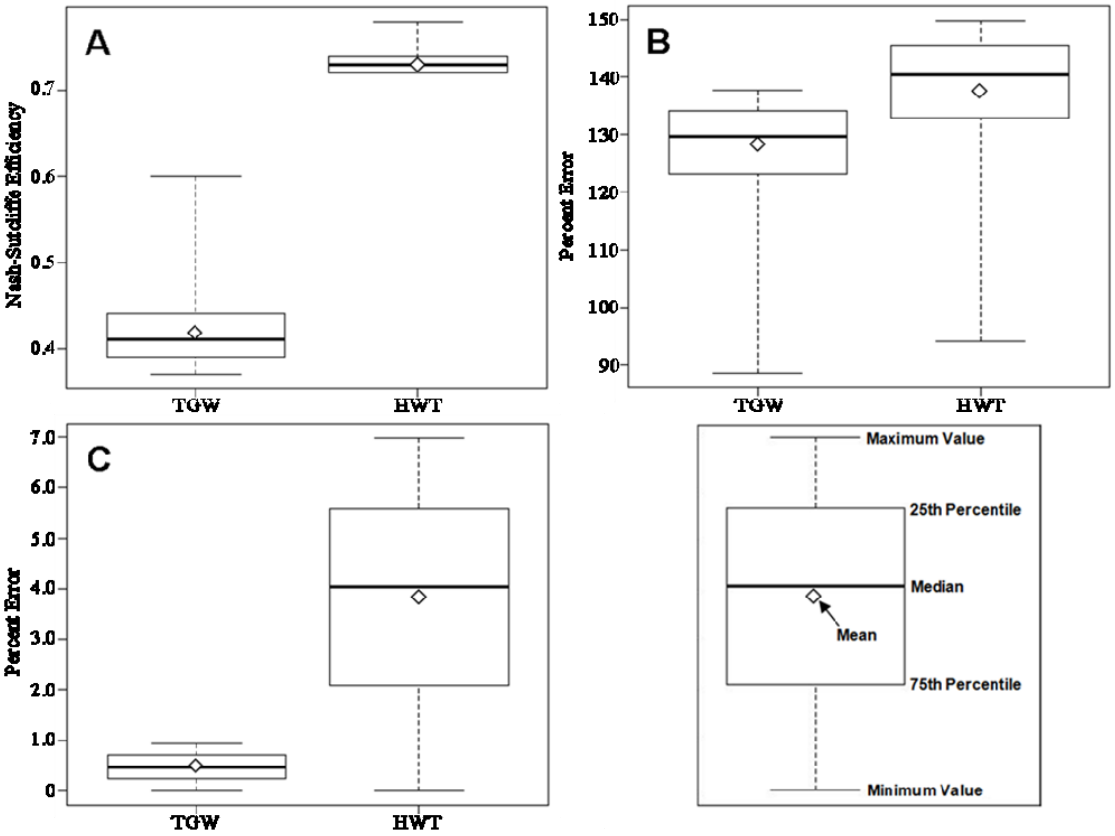

Fig. 7 Boxplots Showing Mean, Median, and Distribution of the TGW and HWT Models When Compared Using Three Metrics: (A) NSE Based on Daily Average Flow Rate, (B) Percent Error for the Lowest 50% of Daily Average Flow Rates (B), and (C) the Percent Error for the Highest 10% of Daily Average Flow Rates

C. Uncertainty Analysis

Model performance variability due to structural and parameter uncertainty was quantified using the ARIL metric and the behavioural model runs for each of the three model performance metrics (Table IV). The ARIL values computed for NSE (average daily flow rate) were calculated over the entire three year simulation period (1461 days). The ARIL values for the lowest 50% and highest 10% lowest daily average flow rates were calculated using only that data from low and high flow events, 731 and 146 days, respectively. The ARIL values for the TGW model were smaller than the HWT model for each of the three performance metrics (Table IV), indicating that the TGW model had less overall uncertainty than the HWT throughout the simulation period including periods of low and high flows. One cause of the greater uncertainty associated with the HWT model is the additional HWT subroutine parameters that are not used in the TGW subroutine (PCW, PGW, UPGW, IFWSC, SRRC, SREXP). These additional parameters and the complexity of the HWT subroutine created more model input and structural uncertainty which increased overall model uncertainty.

TABLE IV ARIL VALUES CALCULATED BASED ON THE THREE PERFORMANCE METRICS

Performance Metric	TGW	HWT
NSE (average daily flow rate)	1.898	2.105
50% Lowest Flows	2.826	3.105
10% Highest Flows	0.366	0.447

D. Parameter Identifiability

Parameter identifiability for the TGW and HWT subroutines was compared by examining the difference between the posterior parameter distributions and the *a priori*, uniform distributions for each parameter. Some parameters showed posterior distributions that were uniform while others showed a significant trend towards a particular parameter value or range. Fig. 8 contains illustrations of selected parameters with posterior distributions where there was a clearly identifiable preference for a particular parameter value or range. Table 5 contains the maximum vertical distance between the *a priori* and posterior parameter distributions; vertical distances ≥ than 0.1 are bolded. The HWT subroutine yielded 20 identifiable parameters across all three model performance metrics, while the TGW subroutine yielded 17. Considering the fact that the HWT subroutine used three more stochastic parameters than the TGW model, 21.5% of the HWT parameters across all three metrics were found to be identifiable compared to 20.2% of the TGW parameters.

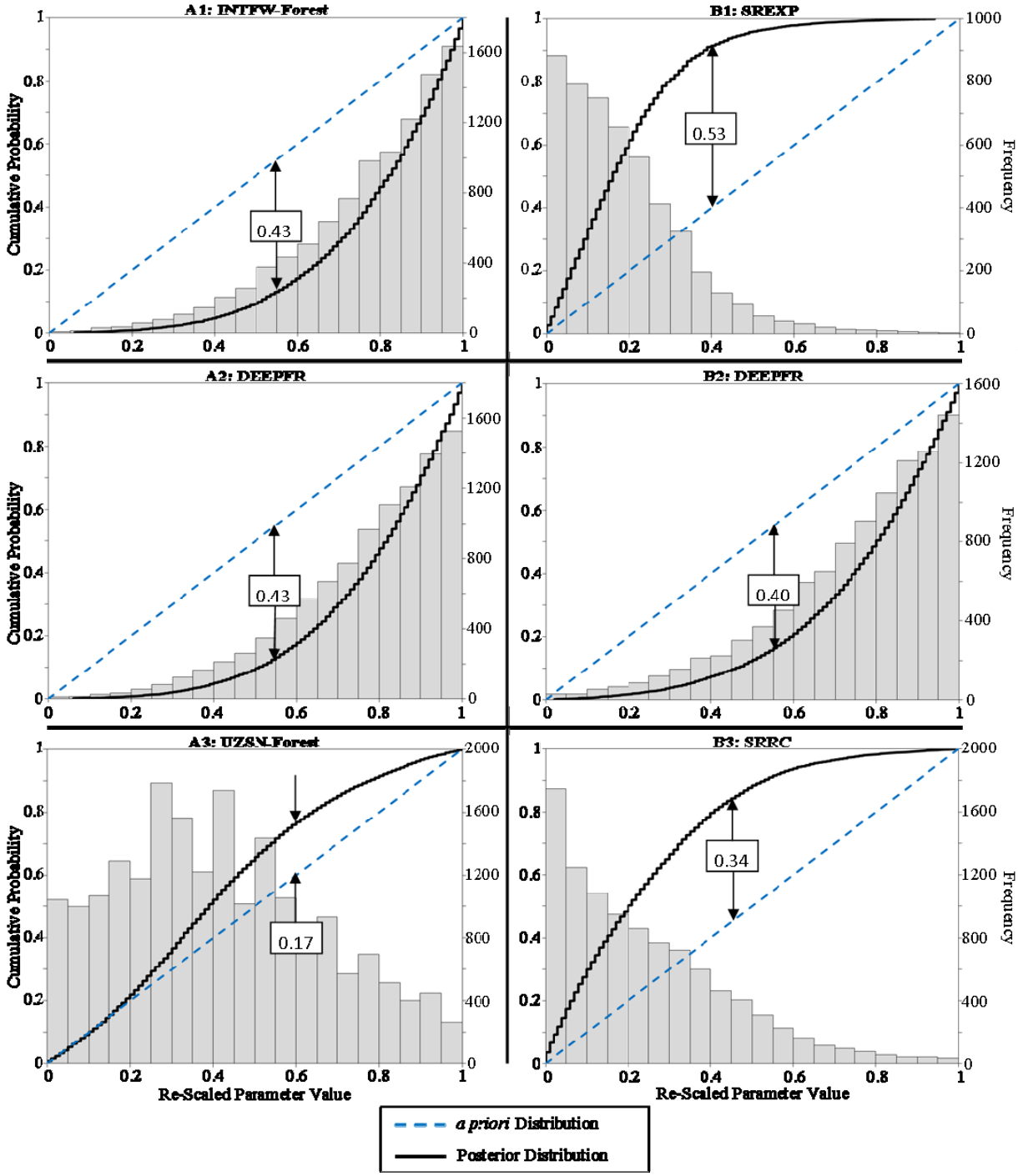

Fig. 8 Illustration of Selected Identifiable Posterior Parameter Distributions for the TGW (A) and HWT (B) Models and the Three Performance Metrics: NSE (1), 50% Lowest Flows (2), and 10% Highest Flows (3)

TABLE V MAXIMUM VERTICAL DISTANCE BETWEEN PARAMETER *A PRIORI* DISTRIBUTIONS AND POSTERIOR DISTRIBUTIONS

Parameter	TGW			HWT		
	NSE	Lowest 50% of Flows	Highest 10% of Flows	NSE	Lowest 50% of Flows	Highest 10% of Flows
AGWETP	0.05	0.19	0.01	0.01	**0.24**	0.02
AGWETP-Wetlands	0.01	0.03	0.02	0.01	0.03	0.01
AGWRC	0.02	0.09	0.03	0.07	**0.35**	**0.12**
BASETP	**0.12**	**0.33**	0.03	0.01	**0.27**	0.04
BASETP-Wetlands	0.03	0.05	0.02	0.02	0.05	0.02
DEEPFR	0.04	**0.43**	0.03	0.08	**0.40**	0.07
DEEPFR-Wetlands	0.02	0.06	0.01	0.02	0.06	0.02
INTFW-Forest	**0.43**	0.07	0.03	0.02	0.04	0.05
INTFW-Cropland	**0.26**	0.02	0.03	0.02	0.03	0.03
INTFW-Pasture	**0.14**	0.02	0.02	0.01	0.02	0.02
INTFW-Residential	0.04	0.02	0.02	0.01	0.01	0.01
INTFW-Wetlands	**0.14**	0.03	0.02	0.01	0.03	0.03
IRC-Forest	0.06	**0.13**	0.05	**0.13**	**0.12**	**0.13**
IRC-Cropland	0.05	0.05	0.03	0.06	0.06	0.06
IRC-Pasture	0.02	0.03	0.02	0.04	0.03	0.03
IRC-Residential	0.01	0.01	0.02	0.01	0.02	0.02
IRC-Wetlands	0.03	0.03	0.02	0.05	0.04	0.03
LZETP-Forest	**0.11**	**0.14**	0.02	0.05	**0.14**	0.04
LZETP-Cropland	0.06	0.08	0.02	0.03	0.08	0.04
LZETP-Pasture	0.07	0.09	0.02	0.03	0.08	0.03
LZETP-Residential	0.03	0.04	0.01	0.02	0.04	0.02
LZETP-Wetlands	0.06	0.06	0.02	0.03	0.05	0.02
LZSN	**0.33**	0.09	0.08	0.06	0.07	**0.16**
UZSN-Forest	**0.33**	**0.14**	**0.17**	**0.27**	**0.18**	**0.29**
UZSN-Cropland	**0.18**	0.08	0.06	0.12	0.08	**0.13**
UZSN-Pasture	**0.12**	0.06	0.04	0.06	0.07	0.07
UZSN-Residential	0.03	0.02	0.02	0.02	0.03	0.02
UZSN-Wetlands	0.08	0.04	0.03	0.06	0.05	0.08
IFWSC	–	–	–	0.01	0.01	0.02
SREXP	–	–	–	**0.53**	0.06	**0.34**
SRRC	–	–	–	**0.22**	**0.10**	**0.40**
Identifiable Parameters	10	6	1	5	8	7

*Numbers shown in bold indicate an identifiable parameter

IV. CONCLUSION

This study compared the performance and uncertainty of two models that implemented two alternative groundwater simulation subroutines available in HSPF. The two models were applied to Ahoskie Creek, a low-gradient, high water table watershed in the Coastal Plain of northeast North Carolina. The high water table (HWT) groundwater subroutine was developed specifically for use when modeling low-gradient, high water table watersheds. This specialized subroutine employs a more sophisticated conceptualization of the groundwater system than HSPF's traditional groundwater subroutine (TGW).

Model performance was assessed using three metrics that compared overall model performance, and the performance of the models when predicting the highest and lowest flows. We used Monte Carlo simulations to generate sufficient model output data to quantify model structure and parameter uncertainty associated with the two groundwater subroutines. Further, we examined *a priori* groundwater subroutine parameter distributions (using the Monte Carlo simulations) with posterior parameter distributions derived from behavioural parameter sets to quantify parameter identifiability, a surrogate measure of the relative ease of model calibration.

In this study, both models performed well when simulating the 10% highest daily average flow rates, falling well within the HSPEXP criteria of +/- 15%. Neither model performed well when simulating the 50% lowest daily average flow rates. Both models greatly exceeded the +/- 10% HSPEXP criterion. The HWT model, however, significantly outperformed the TGW model when simulating daily average flow (NSE) over the full three-year assessment period, an indication that the HWT model out-performed the TGW over the full range of simulated flows.

Given the inherent limitations associated with hydrological models and the heterogeneity of watersheds, various forms of

model uncertainty are unavoidable. Therefore, a balance between model performance and model uncertainty should be considered when selecting an appropriate model for a given application. In this study, based on the Average Relative Interval Length (ARIL) metric which was used to quantify the uncertainty present in the two models resulting from both groundwater subroutine model structure and parameter uncertainty, the HWT model exhibited slightly more combined uncertainty than did the TGW model. However, both the TGW and HWT models exhibited a similar degree of parameter uncertainty, each producing roughly the same proportion of identifiable parameters. Although this study did not directly address the level of effort required to calibrate either the HWT or TGW groundwater subroutine, if one subroutine had yielded a significantly higher number of identifiable parameters, it follows that calibration of that subroutine would likely be more straightforward.

When choosing between the two groundwater subroutines evaluated here, the context in which the HSPF model is being applied must be considered. Although the HWT the model could be considered more accurate, it did produce greater uncertainty than the TGW model. Therefore, in an application where the accuracy of hydrology simulation is perhaps the most critical concern, i.e., an application where accurately predicting pollutant loading is the issue (e.g., TMDL development), the increased uncertainty associated with the HWT model may be tolerable. In the case of the TMDL development application, the greater uncertainty associated with the HWT model could be accommodated with the TMDL margin of safety (MOS). In applications where the balance between model performance and uncertainty is more critical, the choice of which groundwater subroutine to employ is less clear cut.

REFERENCES

[1] V. P. Singh and D. A. Woolhiser. Mathematical Modeling of Watershed Hydrology. Journal of Hydrology 7:270-292, 2002.

[2] K. Beven. Changing Ideas in Hydrology -- The Case of Physically-Based Models. Journal of Hydrology 105:157-172, 1989.

[3] R. B. Grayson, I. D. Moore and T. A. McMahon. Physically Based Hydrologic Modeling: 1. A Terrain-Based Model for Investigative Purposes. Water Resour. Res. 28:2639-2658, 1992.

[4] J. C. Refsgaard. Parameterisation, Calibration and Validation of Distributed Hydrological Models. Journal of Hydrology 198:69-97, 1997.

[5] K. W. Migliaccio and P. Srivastava. Hydrological Components of Watershed-Scale Models. Transactions of the American Society of Agricultural and Biological Engineers 50:1695-1703, 2007.

[6] A. S. Donigian, Jr., B. R. Bicknell and J. C. Imhoff. Hydrological Simulation Program - Fortran (HSPF). In: Computer models of watershed hydrology, V. P. Singh, ed. Water Resources Publications, Colorado, pp. 395-442, 1995.

[7] B. R. Bicknell, J. C. Imhoff, J. L. Kittle, T. H. Jobes and A. S. Donigian. Hydrological Simulation Program - Fortran: HSPF Version 12 User's Manual. U.S. Environmental Protection Agency, Athens, GA, 2001.

[8] P. Duda, J. Kittle, Jr., M. Gray, P. Hummel, R. Dusenbury. WinHSPF version 2.0: An interactive windows interface to HSPF: User manual. Contract No. 68-C-98-010. USEPA, Washington D.C., 2001.

[9] D. K. Borah, D. K. and M. Bera. Watershed-Scale Hydrologic and Nonpoint-Source Pollution Models: Review of Mathematical Bases. Transactions of the American Society of Agricultural and Biological Engineers 46:1553-1566, 2003.

[10] B. L. Benham, C. Baffaut, R. W. Zeckoski, K. R. Mankin, Y. A. Pachepsky, A. M. Sadeghi, K. M. Brannan, M. L. Soupir, and M. J. Habersack. Modeling bacteria fate and transport in watershed to support TMDLs. *Trans. ASABE* 49(4): 987-1002, 2006.

[11] B. L. Benham, , K. M. Brannan, G. R. Yagow, R. W. Zeckoski, T. A. Dillaha, S. Mostaghimi, and J.W. Wynn. Development of bacteria and benthic TMDLs: a case study, Linville Creek, Virginia. *J. Environ. Qual.* 34:1860–1872, 2005.

[12] MapTech. E. coli, Phased Benthic, and Phased Total PCB TMDL Development for Levisa Fork, Slate Creek, and Garden Creek. http://www.deq.virginia.gov/portals/0/DEQ/Water/TMDL/apptmdls/tenbigrvr/levisa.pdf. Accessed April 2013. Virginia Department of Environemntal Quality, Richmond, VA. 165 pp, 2011.

[13] M. N. Beaulac and K. H. Reckhow. An Examination of Land Use - Nutrient Export Relationships. Journal of the American Water Resources Association 18:1013-1024, 1982.

[14] I. Valiela, J. Costa, K. Foreman, J. Teal, B. Howes and D. Aubrey. Transport of Groundwater-Borne Nutrients From Watersheds and Their Effects on Coastal Waters. Biodegradation 10:177-197, 1999.

[15] P. .J. F. Yeh, and E. A. B. Eltahir. Representation of Water Table Dynamics in a Land Surface Scheme. Part I: Model Development. Journal of Climate 18:1861-1880, 2005.

[16] J. Zhang A. Said and M. Ross. Approach Using Active Groundwater Storage for Hydrologic Model Calibration in West-Central Florida. Journal of Irrigation and Drainage Engineering 136:58-62, 2010.

[17] A. Said, M. Ross and K. Trout. Calibration of HSPF Using Active Ground Water Storage. In: World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, K. C. Kabbes (Editor). ASCE, Tampa, Florida, USA, pp. 342-342, 2007.

[18] Sheridan, J. M. Peak Flow Estimates for Coastal Plain Watersheds. Transactions of the American Society of Agricultural and Biological Engineers 45:1319-1326, 2002.

[19] Homer, C. H., J. A. Fry and C. A. Barnes. National Land Cover Database, 2012.

[20] USDA-ARS. The Hydrology and Hydrogeology of Ahoskie Creek Watershed, North Carolina: Data and Analysis. U. S. Department of Agriculture - Agricultural Research Service, Washington, D.C., 1977.

[21] USDA-ARS. The Hydrology and Hydrogeology of Ahoskie Creek Watershed, North Carolina: Data and Analysis. U. S. Department of Agriculture - Agricultural Research Service, Washington, D.C., 1977.

[22] USGS. U. S. Geological Survey Digital Raster Graphics (DRG). www.topomaps.usgs.gov/drg, accessed January 2011.

[23] USGS. U. S. Geological Survey National Elevation Dataset (NED). www.ned.usgs.gov, accessed January 2011.

[24] USGS. U. S. Geological Survey National Land Cover Database (NLCD). www.landcover.usgs.gov, accessed January 2011.

[25] V. T. Chow. Open-Channel Hydraulics. New York, McGraw-Hill, 1959.

[26] USEPA. BASINS Technical Note 6. United States Environmental Protection Agency, Washington D.C., 2000.

[27] S. M. Kim, B. L. Benham, K. M. Brannan, R. W. Zeckoski and J. Doherty. Comparison of Hydrologic Calibration of HSPF Using Automatic and Manual Methods. Water Resoures Research 43:W01402, doi: 01410.01029/02006WR004883. 2007.

[28] N. A. Al-Abed, and H. R. Whiteley. Calibration of the Hydrological Simulation Program Fortran (HSPF) Model Using Automatic Calibration and Geographical Information Systems. Hydrological Processes 16:3169-3188, 2002.

[29] A. S. Donigian. Watershed Model Calibration and Validation: The HSPF Experience. Proceedings of the Water Enviroment Federation 2002:44-73, 2002.

[30] J.E. Nash and J. V. Sutcliffe. River Flow Forecasting Through Conceptual Models Part I - A Discussion of Principles. Journal of Hydrology 10:282-290, 1970.

[31] S. K. Jain and K. P. Sudheer, 2008. Fitting of Hydrologic Models: A Close Look at the Nash--Sutcliffe Index. Journal of Hydrologic Engineering 13:981-986, 2008.

[32] A. M. Lumb, R. B. McCammon and J. John L. Kittle. Users Manual for an Expert System (HSPEXP) for Calibration of the Hydrological Simulation Program - Fortran. U. S. Geological Survey Report 94-4168, Reston, VA, 1994.

[33] R. C. Spear, T. M. Grieb and N. Shang, 1994. Parameter Uncertainty and Interaction in Complex Environmental Models. Water Resources Research 30:3159-3169, 1994.

[34] K. Beven, and A. Binley. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. Hydrological Processes 6:279-298, 1992.

[35] T. Wagener and A. Montanari. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resources Research 47:W06301, 2011.

[36] K. Beven. A Manifesto for the Equifinality Thesis. Journal of Hydrology 320:18-36, 2006.

[37] K. J. Beven. Rainfall-Runoff Modelling - The Primer. Chichester, UK, John Wiley & Sons Lt., 2000.

[38] K. J. Beven, ed. Streamflow Generation Processes. Wallingford, UK, IAHS Press, 2006.

[39] D. Tetzlaff and U. Uhlenbrook. Effects of Spatial Variability of Precipitation for Process-Orientated Hydrological Modeling: Results from Two Nested Catchments. Hydrology and Earth System Sciences Discussions 2:119-154, 2005.

[40] M. Sivapalan. Prediction in Ungauged Basins: A Grand Challenge for Theoretical Hydrology. Hydrological Processes 17:3163-3170, 2003.

[41] X. Jin, C.-Y. Xu, Q. Zhang and V. P. Singh. Parameter and Modeling Uncertainty Simulated by GLUE and a Formal Bayesian Method for a Conceptual Hydrological Model Journal of Hydrology 383:147-155, 2010.

[42] R. C. Spear, T. M. Grieb and N. Shang, 1994. Parameter Uncertainty and Interaction in Complex Environmental Models. Water Resources Research 30:3159-3169, 1994.

[43] T. Wagener. and H. Gupta. Model Identification for Hydrological Forecasting Under Uncertainty. Stochastic Environmental Research and Risk Assessment 19:378-387, 2005.

[44] R. C. Spear and G. M. Hornberger. Eutrophication in Peel Inlet—II. Identification of Critical Uncertainties Via Generalized Sensitivity Analysis. Water Research 14:43-49, 1980.

[45] R. Brun, P. Reichert and H. R. Künsch. Practical Identifiability Analysis of Large Environmental Simulation Models. Water Resoures Research 37:1015-1030, 2001.

[46] M B. Beck. Water Quality Modeling: A Review of the Analysis of Uncertainty. Water Resour. Res. 23:1393-1442, 1987.

[47] N. McIntyre, B. Jackson, A. J. Wade, D. Butterfield and H. S. Wheater, 2005. Sensitivity Analysis of a Catchment-Scale Nitrogen Model. Journal of Hydrology 315:71-92, 2005.