07JEI00085
1726-2135/1684-8799

www.iseis.org/jei

Journal of Environmental Informatics 9(1) 29-40 (2007)

Exploring the Variability in Suspended Sediment Yield Using BASINS-HSPF and Probabilistic Modeling: Implications for Land Use Planning

D. Mitsova-Boneva and X. Wang*

School of Planning, University of Cincinnati, Cincinnati, OH 45221-0016, USA

ABSTRACT. The purpose of this study is to examine the use of Monte Carlo simulation method together with BASINS-HSPF, a deterministic water quality simulation model to investigate the variability of sediment yield from various land cover types in the East Fork Little Miami River watershed, Ohio, USA. The study has two main objectives: first, to obtain interval estimates of suspended sediment yield from urban, agricultural and forest land; and second, to generate probability density and cumulative distribution functions for those estimates. As a result, we will know not only the range of possible values of sediment yield in the watershed but also the probability with which those values are likely to occur. BASINS (Better Assessment Science Integrating Point and Non-Point Sources), a modeling system which provides a Window-based interface for the Hydrologic Simulation Program – Fortran (HSPF), is used in the analysis to simulate streamflow and suspended sediment yield. The daily values of suspended sediment yield from various land cover types generated by BASINS-HSPF are used as input to the Monte Carlo simulation technique that generates probability density and cumulative distribution functions. The distributions are used to draw conclusions about the uncertainty in the hydrologic model predictions in the form of confidence intervals for the predicted sediment yield. The study shows that the width of the confidence interval has a critical importance for the evaluation of the model results.

Keywords: BASINS, HSPF, hydrologic response units, Monte Carlo simulation, probability distributions, sediment yield, streamflow

1. Introduction

Simulation has long been used in environmental analysis for approximate solutions to physical problems in systems that are too complex to be approached analytically (Rubinstein and Melamed, 1998; Law and Kelton, 2000; Crawford-Brown, 2001). A simulation model is considered deterministic when all logical relationships between its components and parameters are controlled by pre-specified algorithms which do not involve responses to random conditions (Rubinstein and Melamed, 1998; Law and Kelton, 2000). These also are called physically-based models. Stochastic or probabilistic simulations allow for the inclusion of random components in the modeling process and the output "is treated as a random variable" itself (Crawford-Brown, 2001). As such, they are interpreted through probability distributions. Such models are called empirical or statistically-based models.

The proper uses of physically-based versus empirical or statistically-based models have been discussed by several researchers (Chow et al., 1988; Cashman, 1997; Huang, 1999; Haan and Skaggs, 2003). Theoretically, distributed or semi-distributed watershed models describing the behavior of hydrologic systems based on a set of mathematical equations provide the most reliable information on the complex physical, chemical and biological processes occurring simultaneously in the watershed at different scales of space and time (Chow et al., 1988; Cashman, 1997).

The practical use of those models, however, raised questions about the variation and uncertainty in parameter estimation and output, as well as possible approaches to address them. In addition, physically-based models often produce point estimates that are fixed for any given set of inputs. Several natural processes, however, such as rainfall amount, intensity, and duration, abstractions, site drainage and erosion are stochastic in nature (Haan et al., 1995; Chow et al., 1988; Haan and Skaggs, 2003; Cashman, 1997). Hence, a set of given input parameters can be better understood through probability distributions. Daily precipitation, for example, which drives many hydrological processes, is often considered a random variable which can be reasonably examined through probability distributions (Chow et al., 1988; Huang, 1999). Issues related to the inherent uncertainty of the modeling process also call for a need for probabilistic approaches (Haan and Skaggs, 2003; Paul et al., 2004). On the other hand, purely statistical methods are useful in making predictions within the range of the dataset or establishing correlations between the variables of interest but they cannot explain causal relationship or provide information on the underlying physical processes. Integrating the strengths of deterministic and probabilistic analyses in examining hydrologic processes contributes to fuller understanding of the dynamics of watershed conditions and provides planners, environmental risk assessors, researchers, and citizens with knowledge on the possible range of outcomes of any planned or proposed land-use/land-cover changes.

* Corresponding author: mitsovdp@email.uc.edu

Since the 1970s, physically-based hydrologic simulation models have increasingly been used to assess the impact of various management practices and evaluate short- and long-term effects of changing hydrologic conditions on water quality. Several such models are currently available. BASINS (Better Assessment Science Integrating Point and Non-Point Sources) consists of four such models and data management tools coupled with a Geographic Information System (GIS) interface (USEPA, 1998). The four models are Hydrologic Simulation Program – Fortran (HSPF) (Donigian et al., 1984; Bicknell et al., 1993; Bicknell et al., 1996), Soil and Water Assessment Tool (SWAT) (Arnold et al., 1994; Arnold et al., 1998), the enhanced stream water quality model QUAL2E 3.2 (Brown and Barnwell, 1987), and Pollutant Loading Program (PLOAD). Borah and Bera (2003) evaluated the mathematical structure and parameterization requirements of eleven watershed based hydrologic and pollutant simulation models and compared their data requirements and applicability. The investigators found that HSPF and SWAT are two of the most widely used and reliable hydrologic simulation models.

Van Liew et al. (2003) used SWAT and HSPF to simulate streamflow in eight nested watersheds in the Washita River Basin in southwestern Oklahoma. The investigators compared the error rate between the observed and simulated flow hydrographs using flow duration curves generated by the two models. They found that both HSPF and SWAT simulated streamflow in ranges close to the monitored values (Van Liew et al., 2003). The authors observed that in some applications HSPF produced better agreement between predicted and observed flow which was explained with the use of the Philip's infiltration equation in HSPF versus the use of the SCS runoff curve number in SWAT (Van Liew et al., 2003).

In another application, Saleh and Du (2004) compared HSPF and SWAT based on the standard deviation and mean error of measured and predicted flow. They used Nash and Sutcliffe (1970) equation to compute model efficiency. The study, conducted in the Upper North Bosque River watershed in central Texas, found higher mean error in SWAT daily flow predictions compared to HSPF. The SCS runoff curve number method (used in SWAT) which accounted only for the total rainfall volumes was considered less accurate than the Philip's infiltration equation applied in HSPF. Saleh and Du (2004) also found that both models simulated sediment yield reasonably well, but SWAT produced better results than HSPF in estimating nutrient concentrations. The authors concluded that HSPF had lower efficiency for nutrient predictions than SWAT due to the fact that it had not been designed to incurporate various agricultural management practices (Saleh and Du, 2004).

Bosch et al. (2003) examined the hydrologic effects of urban development, namely the percent change in estimated flood magnitudes for return periods of 10, 5, 2 and less than one year, and percent change in infiltration which contributes to aquifer recharge. The authors used HSPF to investigate eleven development scenarios, based on high, medium and low density residential development and found that development in headwaters has the highest hydrological impact on the watershed both in terms of increased runoff and decreased infiltration (Bosch et al., 2003). The investigators did not comment on the HSPF performance in the study but mentioned that the constraints of parameterization need to be considered in assessing the validity of results (Bosch et al., 2003).

Although many studies report on the performance of HSPF and other hydrologic and water quality modeling systems, only a few of them explicitly address the issues of uncertainty and variability in outputs. Usually variability is associated with changes in natural conditions while uncertainty stems from the modeling process itself as a result of insufficient knowledge and subjective judgment (Cullen and Frey, 1999; Bates et al., 2003; Paul et al., 2004). Chen et al. (1998) observed four major sources of uncertainty in the modeling process: (i) data deficiencies; (ii) "limited representativeness of point measurements"; (iii) scarcity of on-site data on channel morphology; and (iv) errors generated by the modeling system.

Albek et al. (2004) applied HSPF to investigate how an increase in the mean annual temperature and changes in land cover will affect the discharge rate in the Seydi Suyu watershed in Turkey. The model performance was evaluated using three statistical tests: the correlation coefficient, *t*-test and chi-square test (Albek et al., 2004). All tests were statistically significant which indicates good representation of watershed conditions by the HSPF-simulated results. The simulated daily flows, although showing greater variability, were also thought to be adequate (Albek et al., 2004).

Paul et al. (2004) used first-order approximation to evaluate the impact of five calibration parameters on the variance of HSPF-simulated output in Salado Creek, Texas. The authors observed that even a negligible deviation in accuracy when calibrating these parameters set out a significant error margin in the output (Paul et al., 2004). Im et al. (2004) encountered difficulties in matching sampled data with the daily mean concentrations simulated by HSPF. The discrepancy was attributed to the fact that field data consisted of point measurements while the HSPF predictions were daily averages, and for this reason the HSPF results were considered valid if the field record falled within an interval bounded by the lowest and highest simulation over a 3-day period (Im et al., 2004).

In sum, the literature has raised a number of questions regarding the precision of output results in environmental simulation modeling. More specifically, uncertainty in the modeling results is influenced by: (1) use of different theories and equations; (2) use of different calibration and validation techniques; (3) often lack of sufficient knowledge of specific watershed conditions to allow for accurate model calibration; (4) use of relatively small number of point measurements to calibrate simulated average daily values; (5) application of constant, empirically estimated coefficients for spatial and temporal aggregations; (6) the stochastic nature of some of the inputs; (7) measurement errors; (8) modeling errors; and (9) missing data.

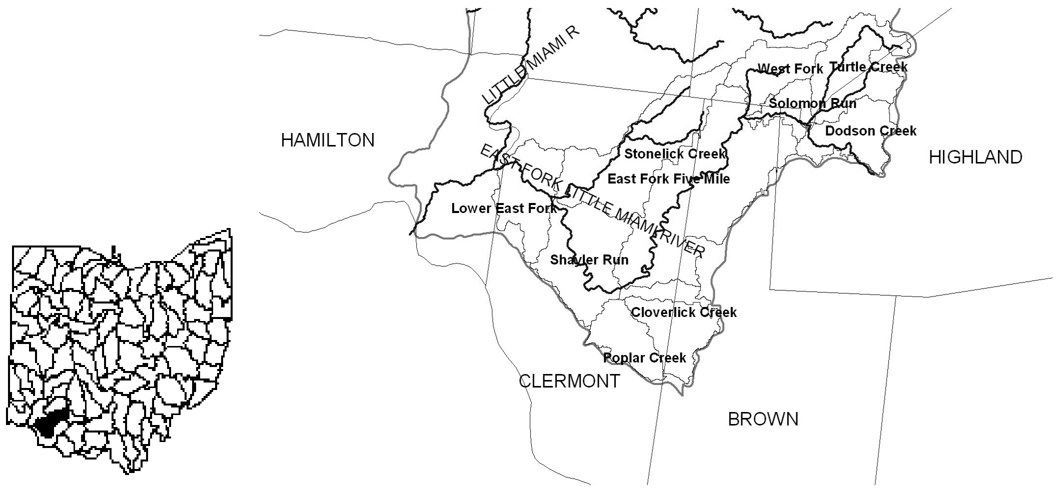

Figure 1. Location of the East Fork Little Miami River watershed in Ohio, USA.

Given these arguments, it is obvious that uncertainty analysis will enhance the validity of the modeling results. Monte Carlo simulation (MCS) has been considered an appropriate technique for conducting probabilistic analyses in many environmental exposure and risk assessment studies. The U.S. Environmental Protection Agency (USEPA) has adopted a policy on the use of Monte Carlo techniques as a means to find multi-scenario solutions to environmental risk problems and examine the sources of variability and uncertainty (USEPA, 1997). The use of Monte Carlo techniques has been prompted by the recognition that in many cases we do not possess full knowledge of the processes underlying an outcome, and even when those processes are well represented in algorithms, the knowledge of theoretical distribution of an outcome could still be rather limited. Even when such knowledge exists, data may not satisfy all assumptions built in the statistical theory. Fan et al. (2002) found that Monte Carlo simulation is particularly relevant in the following situations: (i) the theory does not provide an analytical framework for examining the output distributions; (ii) "the theory about the statistic of interest is weak" (p.12); or (iii) prior knowledge simply does not exist. In these and similar cases, the Monte Carlo method allows for empirical estimation of the sampling distribution characteristics without referring to "the theoretical expectations of those characteristics" (Fan et al., 2002).

Haan et al. (1995) used Monte Carlo simulation to generate cumulative and probability density functions of simulated mean monthly flow. The study investigated the relationship between the modeled flow and the Soil Conservation Service (SCS) runoff curve number CN. The authors found that MCS is a useful technique for estimating uncertainty in inputs and outputs of deterministic modeling. In two studies – on hydrology and the nitrogen cycle, Haan and Skaggs (2003) applied Monte Carlo methods in conjunction with sensitivity analysis and first-order approximation to investigate the expected normal and log-normal distributions for nitrogen loss at the Tidewater Research Station in Plymouth, North Carolina. Korre et al. (2002) used MCS to examine the probability of exceedence of the daily mean concentrations of Pb in the soil substrate around Lavrio, Greece.

Built upon the literature, we use an integrated approach based on coupling BASINS-HSPF and Monte Carlo simulation to study the impact of land cover on sediment yield in the East Fork Little Miami River watershed, Ohio. BASINS-HSPF is used to simulate the production and removal of suspended sediment from pervious and impervious urban surfaces, agricultural, and forest land. The simulated values are entered as input to the Monte Carlo simulation to generate probability density and cumulative distribution functions for the output estimates. More specifically, the study examines the sediment yield values generated by HSPF in terms of probability distributions and derives conclusions about the impact of different land cover types on the likelihood that a certain level of sediment export may occur. The distributions are used to quantify the uncertainty in the model predictions in the form of confidence intervals for the predicted sediment yield. Since both measured data and predictions from deterministic modeling may vary within wide range of values, the method suggested in this study is helpful in approaching questions such as: Which values of suspended sediment yield are most likely to occur? What is the probability that the maximum simulated value will be observed? What is the range of values enfolded between the 0.05 and 0.95 percentiles, or between 0.01 and 0.99 percentiles? In many cases, such information derived from probability density functions is more useful in the decision-making than the point estimates obtained through a deterministic modeling process.

2. Background Information

2.1. Study Area

The study area is the East Fork Little Miami River

(EFLMR) watershed, which covers approximately 506 square miles (1,313 square kilometers) in the southeastern part of the Little Miami River sub-basin (HUC # 05090202) (Figure 1). The river originates north of New Vienna at an elevation of 1,140 feet (348 meters) above the sea level and drops to 492 feet (150 meters) at its confluence with the Little Miami River (CCOEQ, 2000). The average change in longitudinal slope is 7.6 feet per mile (1.4 meters per kilometer). The confluence of the East Fork Little Miami River and the Little Miami River occurs 3.5 miles (5.6 kilometers) south of Milford, just east of the City of Cincinnati. Topographically, the watershed is subdivided into two sub-watersheds. The eastern part of the East Fork Little Miami River watershed is characterized by steeper topography and variations in elevation. The western part is flatter, consisting mainly of gently rolling hills and wider floodplain (CCOEQ, 2000).

Figure 2 shows the land use and soil association maps of the EFLMR watershed. The drainage area of the lower East Fork Little Miami River lies almost entirely in Clermont County, OH, and covers an area of 320 square miles (830 square kilometers). This part of the watershed is a rapidly urbanizing area due to its proximity to Cincinnati. The primary land uses, however, are still agricultural, pasture, forest and low-density residential with the exception of scattered commercial development along highways.

The eastern portion of the watershed is split between Brown, Clinton and Highland counties, Ohio. The headwaters total drainage area is approximately 195 square miles (506 square kilometers) of which 29% fall in Brown, 34% in Clinton and 34% in Highland county (CCOEQ, 2005). Agricultural and forest land covers around 80% of the headwaters. Residential, commercial and industrial development account for less than 20% and are clustered around small communities such as New Vienna, Lynchburg and Fayetteville (MRLC, 1992; CCOEQ, 2000).

The East Fork Little Miami River watershed has temperate climate characterized by cold dry winters and warm humid summer seasons. The average monthly precipitation is 3.5 inches (9 centimeters) (CCPEQ, 2000). The spring and the summer are the wettest seasons with approximately 60% of the total annual precipitation.

The advances and retreats of glaciers during the Illinoisan (130,000 to 300,000 years ago) and Wisconsin (14,000 to 24,000 years ago) ages shaped the landscape and the drainage patterns in the watershed. The parent material below the glacial till and the soil cover consists mainly of shale substrates (CCOEQ, 2000). The dominant soil associations are Clemont-Avonburg-Rossmoyne (OH042) and Rossmoyne-Avonburg-Bonnell (OH051) which account for 54.27 and 32.12% of the soils in the watershed, respectively (71,248 and 42,169 hectares). Rossmoyne-Eden-Cincinnati association (OH052) and Miami-Miamian-Xenia (OH040) cover 3.4% and 3.23% respecttively. Fincastle-Brookston-Miamian (OH038) constitutes 1.15% or 1,510 hectares. Stream network and reservoirs cover 840 hectares or 0.64% of the watershed area. Avonburg series are somewhat poorly drained and exhibit seasonal wetness. Although Cincinnati and Rossmoyne soils are relatively well drained, they contain a fragipan clay layer between loess and glacial till that inhibits downward movement of water and contributes to the formation of perched water tables above it. In sloping landscapes, the lateral subsurface flow that develops above the fragipan layer affects the transport of dissolved and suspended constituents to surface and groundwater (Calmon, 1997; CSWCD, 2002).

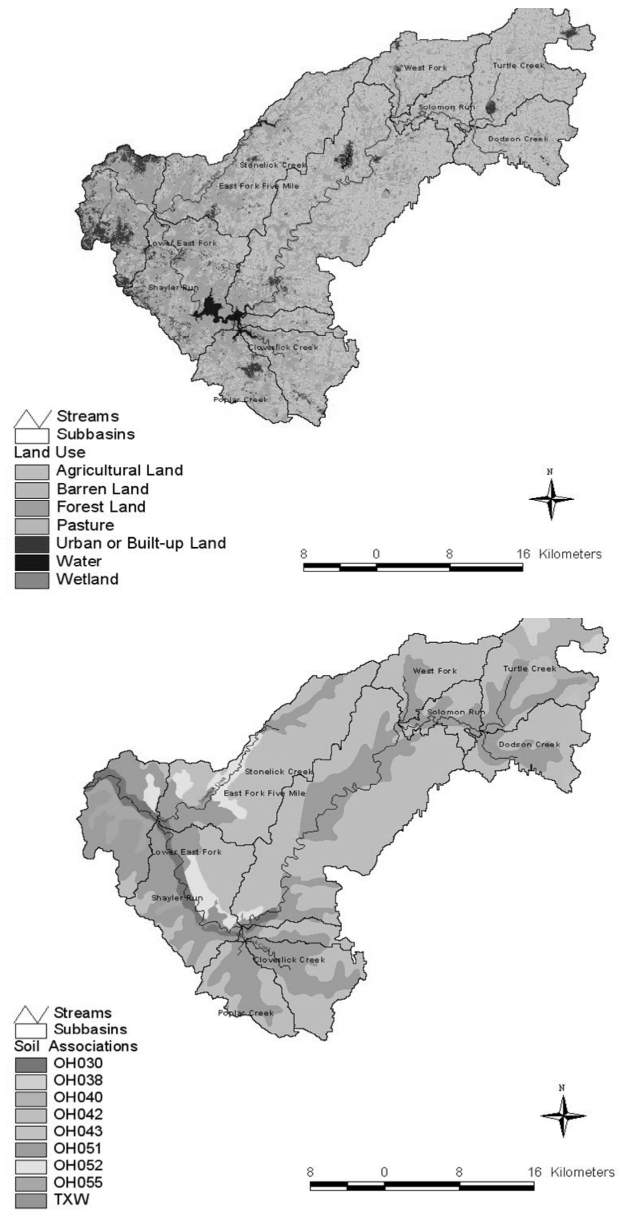

Figure 2. Land use and soil associations maps of the East Fork Little Miami River watershed.

2.2. Data

Data used in the study include three major components: topography and stream network; soils and land cover; and meteorological time series. Digital elevation models (DEM) and elevation grids were obtained through the BASINS Web

Download Tool. In order to be used in hydrological analysis, the DEM data sets were converted to ArcView shapefiles and re-projected from geographic coordinates to a North American Datum of 1983 (NAD83) which uses the Geodetic Reference System of 1980 (GRS80). The DEM shapefile was used with the BASINS Manual Delineation Tool to create a boundary shapefile for the East Fork Little Miami River watershed. The elevation grid file and the National Hydrography Dataset (05090202NHD), which contains detailed information about surface water features, were then used with the BASINS Automatic Delineation Tool to derive the land and stream network segmentation within the watershed.

The Geographic Information Retrieval Analysis System (GIRAS) is a land use data compiled by NASA high-altitude missions in the late 1970s (Saunders and Maidment, 1996). This data file is available through BASINS Web Data Tool. It provides reference to the land use conditions in the 1980s. The GIRAS data were not used in the study. More recent land use data were obtained from the Multi-Resolution Land Characteristics Consortium (MRLC) which provides the 1992 National Land Cover Dataset (NLCD) in grid format. The data were converted to a shapefile. The soils data were obtained through the USEPA website. It is available from the Soil Survey Geographic Database (STATSGO) and is based on the soil association classification. The soil and land use shapefiles were reclassified using the overlay function within BASINS and a soil/land-use grid was created as an input for the HSPF model.

HSPF requires the detailed meteorological time series at hourly time steps to run the simulations. The input time series include hourly precipitation, hourly temperature, hourly dewpoint temperature, cloudiness, wind speed, atmospheric pressure and solar radiation. Potential evapotranspiration is calculated using the Penman equation. BASINS provide the user an opportunity to select from a set of Watershed Data Management (WDM) files which contain data from various weather stations. The WDM file selected for this study contains meteorological data from the Covington WSO Airport Weather Station in Northern Kentucky. Recorded stream flow data from the USGS gauge station at Perintown were added to the project WDM for calibration and validation purposes.

3. Methodologies

The study applies a physically-based hydrologic model, HSPF, in conjunction with a Monte Carlo random sampling technique to examine variability in sediment yield in an urbanizing watershed. The method includes two components: process-oriented physical modeling and probabilistic modeling. It is carried out in five consecutive steps:

(1) Create the hydrologic response units (HRUs) within the EFLMR watershed. HRUs are sub-watershed units that have similar hydrological response characteristics based on topography, slope, and soil.

(2) Simulate the hydrologic response of the watershed with HSPF assigning values to the input parameters based on existing knowledge of the watershed conditions and the literature.

(3) Calibrate and validate the hydrological parameters.

(4) Simulate sediment yield with HSPF.

(5) Generate probability density and cumulative distribution functions for the sediments yields using Monte Carlo random sampling technique.

3.1. Watershed Delineation, Hydrology and Sediment Modeling

Using BASINS manual and automatic delineation tools, eleven hydrologic response units, sufficiently small to respond uniformly to meteorological conditions and accommodate one value for each physical parameter (slope, elevation, soil characteristics, geologic setting, and channel morphology) were identified. HSPF contains three basic modules designed to simulate hydrologic responses of different types of land surfaces: pervious land (PERLND), impervious surface (IMPLND) and in-stream (RCHRES). Table 1 lists the parameters and calibration values used to create a hydrological model of the EFLMR watershed. The initial parameter values were assigned based on knowledge of the existing physical conditions in the watershed. Fifteen parameters were adjustted during the calibration process. The calibration period covered June 1992 through December 1993.

Table 1. HSPF Hydrology Parameters and the Initial Values

Parameter	Initial value	Calibrated value
Overland flow:		
INFILT (mm^{-h})	4.06	10.16
INFEXP	2.00	2.00
UZSN (mm)	28.65	22.86
INFILD (mm)	2.00	2.00
NSUR	0.15	0.10
Subsurface flow:		
INTFW	2.50	3.00
IRC (day^{-1})	0.50	0.65
Groundwater:		
LZSN (mm)	152.4	254
DEEPFR	0.80	0.50
AGWRC (day^{-1})	0.98	0.99
Evapotranspiration:		
CEPSC (mm)	2.54	5.08
LZETP	0.60	0.65
BASETP	0.02	0.01
AGWETP	0.00	0.10

Four parameters, INFILT, INTFW, INFEXP and UZSN, were calibrated to improve the simulation of streamflow. According to the USEPA Technical Note 6 (USEPA 2000), INFILT determines what proportion of the rainfall abstractions will be diverted towards surface runoff and what proportion will be distributed within the soil column. INFILT is an index to the average soil infiltration rate, depending primarily on the soil capacity to absorb moisture. Low values of

INFILT result in higher volumes of surface runoff because they allow the retention of smaller amounts of water in the subsurface storage zone (USEPA, 2000). INFILT is closely related to the upper zone storage parameter UZSN which is influenced by geomorphology and season (USEPA, 2000). Low values of UZSN indicate less storage in the upper subsurface zone, and therefore, increased surface runoff. INFEXP specifies the infiltration equation exponent (USEPA, 2000). It is set to 2.0 by default and needs to be adjusted only under specific watershed conditions. NSUR stands for Manning's roughness coefficient for overland flow on impervious surfaces and measures the friction between the water and land surface in terms of hydraulic efficiency (Chow et al., 1988). The lower the values of this parameter, the higher the potential for increased velocity of flow and larger runoff volumes. INTFW, the interflow-inflow coefficient, controls the distribution of moisture between short-term detention storage and interflow. It determines how much water will become interflow and has no effect on the total volume of runoff, but higher INTFW values will lower peak flows (USEPA, 2000). The interflow recession coefficient IRC is computed as a ratio of the daily interflow volumes in two consecutive days and shows the rate at which recession occurs. It determines when the peak flow will become a baseflow (USEPA, 2000). In HSPF, four parameters determine hydraulic behavior in the saturated zone. The lower zone nominal storage (LZSN) is a function of rainfall intensity and duration, and permeability (USEPA, 2000). It determines the amount of water available for aquifer recharge. DEEPFR accounts for abstractions available for groundwater storage. The groundwater recession coefficient (AGWRC) is calculated as a ratio of groundwater discharge volumes in two consecutive days. Values in the range of 0.98 to 0.996 are usually assigned to this parameter (Laroche et al., 1996; Saleh and Du, 2004; Albek et al., 2004; Van Liew et al., 2003). A value of 0.99 has been used in this study.

Four evapotranspiration parameters were examined and calibrated. The interception storage capacity coefficient (CEP SC) describes abstractions from rainfall resulting from retention by plant leaves. Dense forest cover is capable of retaining significant amount of water (USEPA, 2000). Given the land cover in the EFLMR watershed is mostly cropland and scattered forest cover, a value of 0.15 was assigned to this parameter. HSPF also accounts for evapotranspiration in the upper and lower storage zones through parameters LZETP and BA SETP. Those parameters were adjusted to improve the low flow estimation. LZETP was set to 0.65 to account for the predominant type of vegetation, mostly raw crops and forest. BASETP which indicates the presence of riparian vegetation was set to 0.018. The model has been validated for the period of September 1994 through December 1995.

Sediment parameters in HSPF were calibrated based on the target loads calculated using RUSLE2 (Revised Universal Soil Loss Equation 2) (Yoder et al., 2005). TAUCS and TAU CD (the parameters for critical bed shear stress for deposition and scour respectively) were calibrated to improve the simulation of cohesive sediments associated with the predominant soil type in the watershed: silt and clay. Non-cohesive sediments were not calibrated since the presence of sandy soils is negligible. KSER determines the erosion potential for different types of land uses. A value of 0.15 was assigned for urban land, 0.30 for forest, 0.95 for agriculture, and 0.25 for barren land. Default values for KRER (the coefficient in the soil detachment equation) and JSER (the exponent of the detached sediment wash-off equation) were used. Summary of sediment calibration is presented in Table 2.

Table 2. HSPF Sediment Parameters, Initial and Calibrated Values

Parameter	Initial value	Calibrated value
KRER	0.14	0.35
KSER	0.10	2.50 – 50
AFFIX	0.10	0.002 – 0.008
COVER	0.00	0.50 – 0.97
KEIM	0.10	2.50
TAUCS	0.10	0.15
TAUCD	0.50	0.80

A comparison between the HSPF-estimated annual sediment yield from agricultural, forest, and urban land and the RUSLE2-predicted erosion rates indicated that HSPF performed satisfactorily in estimating sediment yield. RUSLE2 is an empirical model developed by the U.S. Department of Agriculture – Agricultural Research Service which predicts net annual erosion based on soils, erodibility, slope, and vegetation (Yoder et al., 2005). It is not designed to simulate sediment transport, and therefore can only provide rough estimates of the expected sediment yield from various land cover categories. Tables 3 and 4 show acceptable agreement between annual expected erosion rates as calculated with RUSLE2 and the HSPF-simulated values (Table 3). Results show that HSPF slightly overestimated the erosion rates for forest land cover.

Table 3. HSPF-estimated Annual Sediment Yield from Various Land Cover Types (ton/ac)

Year	Forest	Pervious Urban	Agricultural	Impervious Urban
1992	0.254	0.459	0.905	0.0907
1993	0.356	0.356	0.803	0.0943
1994	0.257	0.447	0.997	0.103
1995	0.222	0.485	1.000	0.109

3.2. Monte Carlo Simulation

Two SAS functions (RANNOR and RANTBL) were used for the random samplings. RANNOR draws random numbers from a log-normal distribution. RANTBL is applied when there is no prior knowledge of the theoretical distribution of the variable "but a stepwise approximation of it is available" (Fan et al., 2002). Both functions were applied to determine the probability density and cumulative distribution functions of the HSPF-simulated values. A probability density function (PDF) is a graphical expression of "the likelihood with which values of an input may be obtained" (Cullen and Frey, 1999). A cumulative distribution function (CDF) displays the values of the random variable that are associated with any given percentile of interest, $F(x) = Prob\ (X \leq x)$ (USEPA, 1997). CDF results from summation across the PDF. The x-axis displays the values of the parameter of interest and the y-axis reveal the range within which those values are contained (Cullen and Frey, 1999). The SAS code for the Monte Carlo simulations was adapted from Fan et al. (2002)[1].

Table 4. RUSLE2-calculated Annual Sediment Target Loads for Various Land Cover Types (ton/ac).

Type of land cover	Target loads
Forest	0.203
Urban	0.542
Agricultural	1.172

In Monte Carlo simulation (MCS) the desired parameter θ with a probability distribution $U(x)$ and an integral function $g(x)$ is denoted in the form of (Yakovitz et al. 1978, Law and Kelton 2000):

$$\theta = \int g(x) dU(x) \tag{1}$$

Using random sampling technique, U-distributed random samples $X_1, X_2, X_3, \ldots, X_n$ are generated where X is a continuous random variable. The parameter of interest (θ) is then approximated by:

$$\theta_n = 1/n \sum_{i=1}^{n} g(X_i) \tag{2}$$

If Y is a random variable $c \cdot g(x)$, its expected value $E(Y) = \theta$ will then be estimated by:

$$E(Y) = 1/n \sum_{i=1}^{n} Y_i = c/n \sum_{i=1}^{n} g(X_i) \tag{3}$$

Yakowitz et al. (1978) showed that if the variance $Var(Y)$ is finite, then:

$$E[(\theta_n - \theta)^2] = Var(Y)/n \tag{4}$$

Therefore, $E(Y)$ is an unbiased estimator of θ for sufficiently large n (Yakowitz et al., 1978; Law and Kelton, 2000).

[1] Fan et al. (2002) developed the SAS program for Monte Carlo analysis of bond prices on financial markets.

4. Results and Discussions

Recorded streamflow data from the nearest USGS station (site # 03247500) at Perintown, OH, were compared to the flow modeled with BASINS-HSPF over 18-month calibration period - from June 1, 1992 through April 30, 1993. The model was validated over a 16-month period – from September 1, 1994 through December 31, 1995. The initial simulation of flow underestimated the yearly outflow by -23.8% for 1992, and -18.74% for 1993. After calibration, the error was around 15%. For the validation period the simulated annual discharge was 3.2% below the monitored yearly discharge for 1994 and 7.8% below the monitored value for 1995. Figures 3 and 4 present the results of the hydrologic modeling.

The time series describing the contribution of different land cover types to simulated sediment yield, generated by HSPF, exhibited significant variation. In all four years included in the analysis, values for sediment yield from forest land during the months of January and February were between 0.05 and 21.5 kilograms per acre per day. For agricultural land sediment export during the same period varied from 0.5 to 92 kilograms per acre per day. Low values for sediment export from the watershed were also observed during the month of October. They also coincide with the lowest observed daily precipitation.

The peaks in sediment yield from agricultural and forest land were found in the months of April and July in all four years. The maximum simulated values were 87 kilograms per acre per day for forest land and 234 kilograms per acre per day for agricultural land. The extreme values occurred after major storm events. The simulated values for impervious urban surfaces exhibit less seasonal variation. Maximum sediment daily yield from urban surfaces (pervious and impervious) was approximately 170 kilograms per acre per day. Sediment was also produced between storms on agricultural land during the growing season which coincides with the irrigation practices.

Results indicated that HSPF-simulated sediment yield values for any given land cover type vary significantly. By simply examining the time series and the descriptive statistics of the simulated results it was not possible to draw conclusions about the nature of this variation. For this purpose, probability density and cumulative distribution functions of the HSPF estimates were generated using Monte Carlo simulation. The SAS program for the probabilistic analysis first read the HSPF generated values and performed normality test. Since the normal probability plots and both the Shapiro-Wilk and Kolmogorov-Smirnov tests showed departure from normality, the program was required to compute the logarithm of each observation, and to perform normality tests again.

The fit of the theoretical and the Monte Carlo generated distribution functions was also examined with Kolmogorov-Smirnov goodness-of-fit test (Table 5). The results indicated approximate agreement with the log-normality assumption.

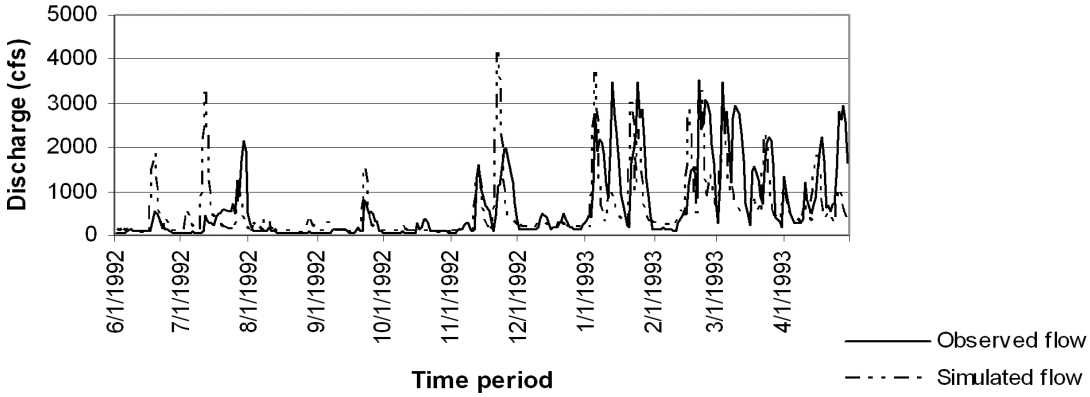

Figure 3. Precipitation, recorded and HSPF simulated flow at the East Fork Little Miami River watershed, OH for the calibration period – June 1992 through December 1995.

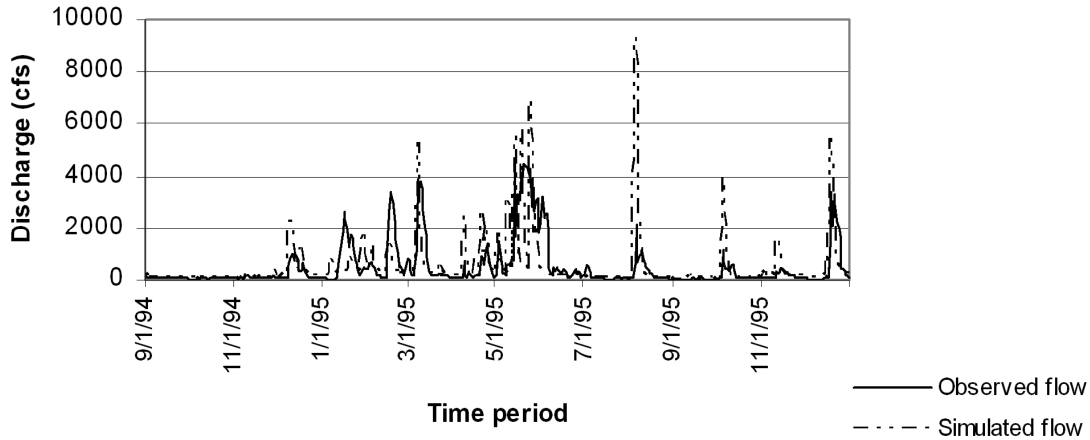

Figure 4. Precipitation, recorded and HSPF simulated flow at the East Fork Little Miami River Watershed, OH, for the validation period – September 1994 through December 1995.

Figures 5, 7 and 9 indicate that the theoretical PDF fits reasonably the randomly generated probability density functions of sediment yield from agricultural and forest land, and and urban (pervious and impervious) surfaces. Fundamental requirement for the Monte Carlo simulation is that the randomly generated sample follows some type of theoretical distribution (Haan and Skaggs, 2003; Fan et al., 2002). Log-normal distributions are commonly used in engineering and natural sciences because they represent physical properties and processes reasonably well (Cullen and Frey, 1999). Log-normal distributions are useful in examining concentrations data because they are based on (0, ∞), or "non-negative values", fit rate-based processes acceptably and allow scientific reasoning with regard to "large asymmetric uncertainties" (Cullen and Frey, 1999).

Random numbers with a theoretical log-normal distribution were generated using as an input the log mean and the log standard deviation of the HSPF-simulated values for each land use type. Numerical stability for the PDF was achieved at 5,000 simulations. However, to ensure constant values 10,000 simulations were run. For the CDF, numerical stability was achieved at 2,000 simulations. Nevertheless, 5,000 simulations were run for the analysis. The input data for the Monte Carlo simulation are summarized in Table 6.

Table 5. Results of the Kolmogorov-Smitnov Goodness-of-fit Test.

Land use type	Test-statistic D	Critical D $\alpha = 0.05$	Critical D $\alpha = 0.01$
Agricultural	0.093	0.035	0.042
Forest	0.056	0.035	0.042
Impervious	0.039	0.035	0.042

Figures 5 through 10 display the probability density and cumulative distribution functions generated with the Monte Carlo technique for the three land use types investigated in the study. The distributions are used to quantify the uncertainty in the model predictions in the form of confidence intervals for the predicted sediment loads to the watershed. The width of the confidence has a critical importance for the evaluation of the model results. It is assumed that a wide confidence interval usually indicates a higher level of uncertainty about the true value of the parameter of interest (Haan and Skaggs, 2003). It is also argued that narrowing the interval will decrease the uncertainty but the likelihood that the model will accurately predict exceedence values also decreases. Intervals that are too narrow, however, may not contain the true value of the parameter of interest (Haan and Skaggs, 2003).

Table 6. Input Data for the Monte Carlo Simulation

Land use type	Log mean	Log standard deviation
Agricultural	4.2112	0.8676
Forest	4.1098	0.8239
Urban impervious	3.1128	0.6770
Urban pervious	3.6345	0.7654

Figures 5, 7 and 9 show the probability density functions of the HSPF-simulated values for sediment yield from various land cover types. They provide indication on the likelihood that a particular HSPF-simulated value will occur. Figures 6, 8 and 10 which display the cumulative distribution functions for sediment yield from forest, agricultural and urban land allows to determine the percentile associated with that particular value. For forest land, for example, we can assume with 95% confidence by visually examining graphs 5 and 6 that values in the range of 2 to 25 kilograms per acre per day are most likely to occur. Therefore, for forest land the likelihood that values above 40 kilograms per acre per day will occur is very low (Figure 6).

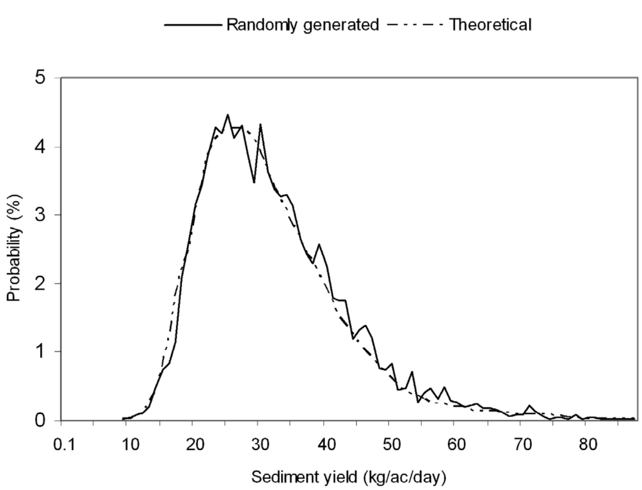

Figure 5. Probability density function for simulated sediment yield from forest land (kg $ac^{-1}day^{-1}$).

Figures 7 and 8 indicate that there is 95% chance that the true value of the daily average sediment yield from agricultural land will be between 20 and 180 kilograms per acre per day. The median value is 78.5 kilograms per acre per day. The analysis indicates that values as high as 234 kilograms per acre per day as obtained with HSPF can be considered extreme events. For urban pervious and impervious surfaces, the PDF and the CDF (Figures 9 and 10) show with 95% confidence that the true value of the average daily sediment yield falls between 2 and 120 kilograms per acre per day. Probability distributions are also very useful in examining various scenarios. For example, if one needs to know the probability that urban surfaces will produce on average 80 kilograms of sediments per acre per day, the CDF on Figure 10 shows that this likelihood is approximately 75%. The probability that the same amount of sediment will be produced by agricultural land is 90%.

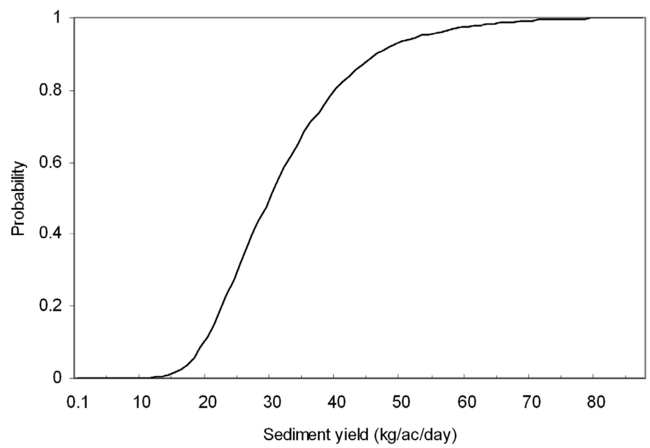

Figure 6. Cumulative distribution function for simulated sediment yield from forest land (kg $ac^{-1}day^{-1}$).

In order to validate the results, the Delta method for confidence interval estimation for lognormal distributions was used to estimate the confidence intervals for the Monte Carlo simulated values. The formula for the η-th quantile estimation of lognormal distribution is given as:

$$\hat{r}_\eta = \exp(\hat{\mu} + \hat{\sigma}(a_\eta)) \tag{5}$$

where $a_\eta = \phi^{-1}(\eta)$ and $\phi^{-1}(\eta)$ is the inverse function of the standard normal CDF (Hahn and Meeker, 1991; Meeker and Escobar, 1998). Table 9 summarizes the confidence intervals for the values simulated with the Monte Carlo technique.

It is useful to examine those probability distributions in terms of whether the observed values are enclosed within the boundaries of the stated intervals as well. It is assumed that if the observed values are bound by the stated interval, then we can claim, for example, that the model is valid in terms of estimating the true value of the parameter of interest with 95% confidence. This assumption, however, has two limitations. First, monitored values are point measurements taken "at a specific point in space and time" while the values generated by HSPF are daily averages (Im et al., 2004). Second, we could not obtain observed values of suspended sediment by land use type in order to compare them with our results. Theoretically, where there is no available data and the true parameter value is unknown, probability and cumulative density functions can still be very helpful in estimating a set of possible outcomes. It should be kept in mind, however, that confidence intervals that are statistically satisfactory may still "render the model predictions too uncertain for the desired application" (Haan et al., 1995). Haan and Skaggs (2003) argue that the uncertainty analysis is not intended to replicate field results; it is rather a process of approximation which allows evaluation of validity of predictions.

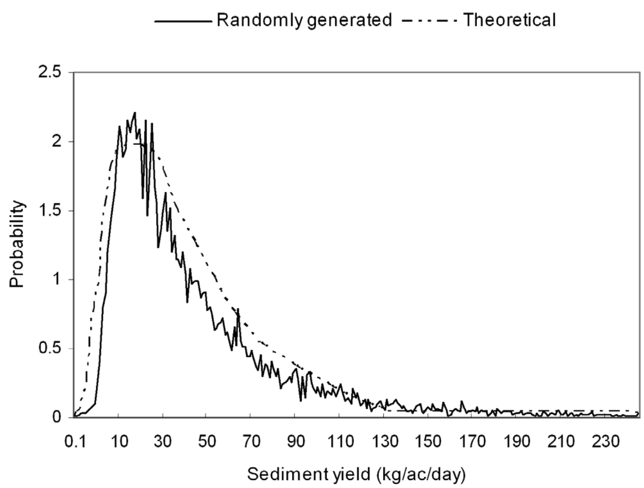

Figure 7. Probability density function for simulated sediment yield from agricultural land (kg $ac^{-1}day^{-1}$).

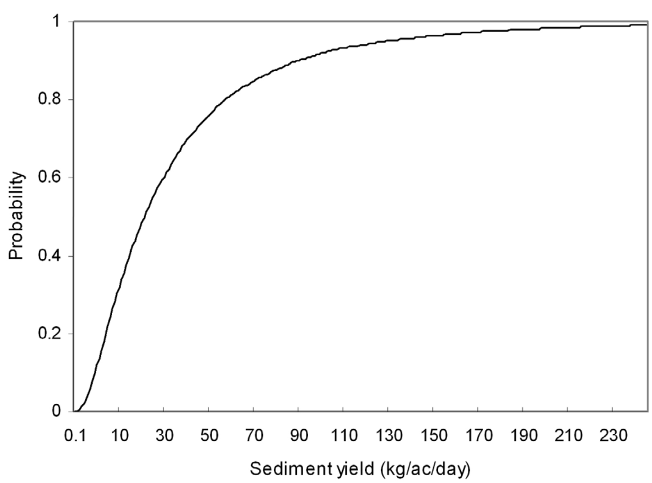

Figure 8. Cumulative distribution function for simulated sediment yield from agricultural land (kg $ac^{-1}day^{-1}$).

4. Conclusions

The study shows that integrating deterministic modeling and probabilistic approaches in examining the relationship between land use and water quality is beneficial and can improve our understanding of the watershed responses to various conditions. BASINS-HSPF was efficient in creating a hydrologic model of the East Fork Little Miami River watershed. The Monte Carlo technique was helpful in examining the HS PF results in terms of uncertainty and variability. It helped to determine the probability that the true value of the parameter of interest would be within a certain range of HSPF simulated values. More specifically, the study focused on estimating confidence intervals of sediment yield predictions from various land cover types using Monte Carlo simulation. It has been found that the log-transformation of the HSPF-simulated data yields confidence intervals that are too narrow for useful applications. The Monte Carlo simulation yielded wider confidence intervals which, given the variation in the datasets, increased the confidence that the true values lies within the stated bounds.

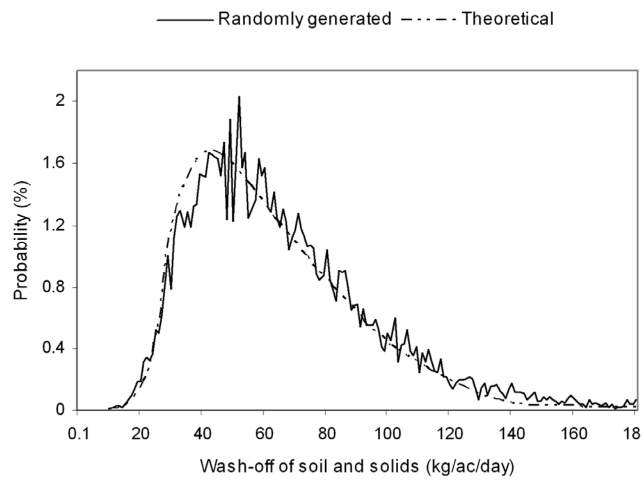

Figure 9. Probability density function of wash-off of soil and solids from urban surfaces (pervious and impervious) (kg $ac^{-1}day^{-1}$).

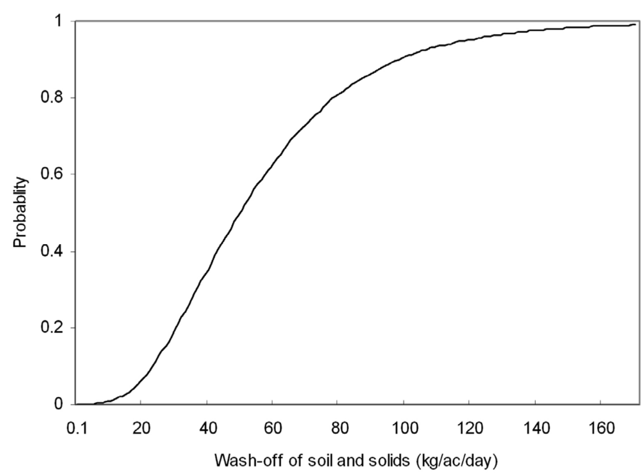

Figure 10. Cumulative distribution function of wash-off of soil and solids from urban surfaces (pervious and impervious) (kg $ac^{-1}day^{-1}$).

The probability density and cumulative distribution functions of simulated values obtained through HSPF and other hydrologic models can play significant roles in decision-making concerning future land uses. The integrated approach presented here provides planners and watershed managers with a tool to combine hydrologic and statistical modeling to investigate issues of concern in land use planning at watershed level.

Table 7. Estimating Confidence Intervals for Monte Carlo Simulated Values

Land use type	η quantile	α_η for lognormal distribution	Confidence interval kg ac^{-1}day^{-1}	Median kg ac^{-1}day^{-1}
Agri.	η = 0.10	-1.282	21.71 – 88.11	56.71
	η = 0.01	-2.362	14.46 – 103.75	
Forest	η = 0.10	-1.282	7.85 - 45.42	28.62
	η = 0.01	-2.362	4.71 - 62.76	
Urban	η = 0.10	-1.282	8.06 - 69.27	43.40
	η = 0.01	-2.362	5.62 - 83.80	

References

Al-Abed, N.A. and Whiteley, H.R. (2002). Calibration of the hydrological simulation program-FORTRAN (HSPF) model using automatic calibration and geographical information systems. *Hydrol. Process.*, 10(10), 3068 -3188.

Albek, M., Ogutveren, U.B. and Albek, E. (2004). Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF. *J. Hydrol.*, 285, 260-271.

Arnold, J.G., Williams, J.R., Srivasian, R., King, K.W. and Griggs, R.H. (1994). *SWAT-Soil and Water Assessment Tool-user Manual*, Agricultural Research Service, US Department of Agriculture, Soil and Water Research Lab.

Arnold, J.G., Srivasian, R., Muttiah, R.S. and Williams, J.R. (1998). Large-area hydrologic modeling and assessment: Part 1. Model development. *J. Am. Water Res. Assoc.*, 34(1), 73-89.

Bates, S., Cullen, A. and Raftery, A. (2003). Bayesian uncertainty assessment in multi-compartment deterministic simulation models for environmental risk assessment. *Environmetrics*, 14, 355-371.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S.Jr. and Johanson, R.C. (1993). *Hydrologic Simulation Program-FOR TRAN* (*HSPF*), User's Manual for Release 10, EPA/600/R-93 /174, USEPA Environmental Research Lab, Athens, GA.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S.Jr. and Johanson, R.C. (1998). *Hydrologic Simulation Program-FOR TRAN* (*HSPF*), User's Manual for Release 11, USEPA Environmental Research Lab, Athens, GA.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jobes, T.H. and Donigian, A.S. (2000). *Hydrologic Simulation Program-FORTRAN* (*HS PF*), User's Manual for Release 12, USEPA Environmental Research Lab, Athens, GA.

Borah, D. and Bera, M. (2003). Watershed-scale hydrologic and non-point-source pollution models: review of mathematical bases. *Trans. ASAE*, 46(6), 1553-1566.

Borah, D. and Bera, M. (2004). Watershed-scale hydrologic and nonpoint-source pollution models: review of applications. *Trans. ASAE*, 47(3), 789-803.

Bosch, D.J., Lohani, V.K., Dymond, R.L., Kibler, D. and Stephenson, K. (2003). Hydrological and fiscal impacts of residential development: Virginia case study. *J Water Res. Plan. Manage.*, 129(2), 107-114.

Brown, L.C. and Barnell, T.O. (1987). *The Enhanced Water Quality Model QUAL2E and QUAL2E-UNCAS Documentation and User Manual*, EPA060003-87-007, USEPA Environmental Research Lab, Athens, GA.

Calmon, M.A. (1997). *Impact of Fragipans on Hydrodynamics and Soil Morphology in Glacial Till Soils in Northeastern Pennsylvania*, Ph.D. Dissertation, Pennsylvania State University, University Park, PA.

Cashman, E. (1997). *Sediment Runoff from Agricultural Watersheds, Variability in Suspended Sediment Transport due to Rainfall*, Ph.D. Dissertation, University Wisconsin-Madison, Madison, WI.

Clermont County Office of Environmental Quality (2005). Lower East Fork Watershed Action Plan-Final Version. http://www.oeq.net/default.php?section=wataction.

Clermont County Office of Environmental Quality (2005). Headwaters Watershed Action Plan-October, 2005 Draft. http://www.oeq.net/default.php?section=wataction.

Clermont Soil and Water Conservation District (2002). Soil Survey of Clermont County, OH. http://www.clermontswcd.org/soilpage.html.

Chen, C., Herr, J.W., Goldstein, R.A., Ice, G. and Cundy, T. (2005). Retrospective comparison of watershed analysis risk management framework and hydrological simulation program-FOR TRAN: Applications to Mica Creek Watershed. *J. Environ. Eng.*, 131(9), 1277-1287.

Chen, Y.D., McCatcheon, S.C., Norton, D.J. and Nutter, W.L. (1998). Stream temperature simulation of forested riparian areas: II. Model application. *J. Environ. Eng.*, 124 (4), 316-328.

Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). *Applied Hydrology*, McGraw-Hill, Boston, MA.

Coffey, M., Workman, S., Taraba, J. and Fogle, A. (2004). Statistical procedures for evaluating daily and monthly hydrologic model predictions. *Trans. ASAE*, 47(1), 59-68.

Crawford-Brown, D.J. (2001). *Mathematical Methods in Environmental Risk Modeling*, Kluwer Academic Publishers, Boston, MA.

Cullen, A. and Frey, H.C. (1999). *Probabilistic Techniques in Exposure Assessment: A Handbook for Addressing Variability and Uncertainty in Models and Inputs*, Plenum Press, New York.

Donigian, A.S.Jr., Imhoff, J.C., Bicknell, B.R. and Kittle, J.L.Jr. (1984). *Application Guide for the Hydrological Simulation Program-FORTRAN*, USEPA Environmental Research Lab, Athens, GA.

Fan, X., Felsovalyi, A., Sivo, S. and Keenan, S.C. (2002). *SAS for Monte Carlo Studies: A Guide for Quantitative Researchers*, SAS Institute, Cary, NC.

Haan, C.T. (2002). *Statistical Methods in Hydrology*, 2nd Edition, Iowa State University Press, Ames, IA.

Haan, C.T., Allred, B., Storm, D.E., Sabbagh, G.J. and Prabhu, S. (1995). Statistical procedures for evaluating hydrologic/ water quality models. *Trans. ASAE*, 38(3), 725-733.

Haan, P. and Skaggs, R.W. (2003). Effect of parameter uncertainty on DRAINMOD predictions: I. Hydrology and yield. *Trans. ASAE*, 46(4), 1061-1067.

Haan, P. and Skaggs, R.W. (2003). Effect of parameter uncertainty on DRAINMOD predictions: I. Nitrogen loss. *Trans. ASAE*, 46(4), 1069-1075.

Hahn, G.J. and Meeker, W.Q. (1991). *Statistical Intervals: A Guide for Practitioners*, John Wiley & Sons, New York.

Huang, Y. (1999). *Spatial and Temporal Variation in Pollutant Concentrations: Implications for Exposure and Risk Assessment of Airborne Pollutants*, Ph.D. Dissertation, University of Michigan, Ann Arbor, MI.

Im, S., Brannan, K.M., Mostaghimi, S. and Cho, J. (2004). Simulating fecal coliform bacteria loading from urbanizing watershed. *J. Environ. Sci. Health*, 39(3), 663-679.

Johnson, M.S., Coon, W.F., Mehta, V.K., Steenhuis, T.S., Brooks, E.S. and Boll, J. (2003). Application of two hydrologic models with different runoff mechanisms to a hillslope of HSPF and SMR. *J. Hydrol.*, 284:57-76.

Korre, A., Durucan, S. and Koutroumani, A. (2002). Quantitative spatial assessment of the risks associated with high Pb loads in soils around Lavrio, Greece. *Appl. Geochem.*, 17, 1029-1045.

Laroche, A.M., Gallichard, J., Lagace, R. and Pesant, A. (1998). Simulating atrazine transport with HSPF in an agricultural watershed. *J. Environ. Eng.*, 122(7), 622-630.

Law, A. and Kelton, W.D. (2000). *Simulation Modeling and Analysis*, 3rd Edition, McGraw-Hill, New York, NY.

Multi-Resolution Land Characteristics Consortium (MRLC). http://gisdata.usgs.net/website/MRLC/viewer.php.

Nash, J.E. and Sutcliffe, J.E. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, 10, 282-290.

Paul, S., Haan, P.K., Mukhtar, S., Metlock, M. and Pillai, S.D. (2004). Analysis of the HSPF water quality parameter uncertainty in predicting peak in-stream fecal coliform concentrations. *Trans. ASAE*, 47(1), 69-78.

Poulter, S. (1988). Monte Carlo simulation in environmental risk assessment-science, policy and legal issues. *Risk, Health, Saf. Environ.*, 7(4), 8-21.

Randolph, J. (2004). *Environmental Land Use Planning and Management*, Island Press, Washington, DC,.

Rubinstein, R. and Melamed, B. (1998). *Modern Simulation and Modeling*, John Wiley & Sons, New York, NY.

Saleh, A. and Du, B. (2004). Evaluation of SWAT and HSPF within BASINS program for the Upper North Bosque River watershed in Central Texas. *Trans. ASAE*, 47(4), 1039-1049.

Saunders, W.K. and Maidment D.R. (1996). A GIS Assessment of Nonpoint Source Pollution in the San Antonio Coastal Basin, CRWR Online Report 96-1, Center for Research in Water Resources, Austin, TX. http://www.ce.utexas.edu/ centers/crwr/reports/online.html.

Tong, S. and Chen W. (2002). Modeling the relationship between land use and surface water quality. *J. Environ. Manage.*, 66, 377-393.

US Environmental Protection Agency (2000). *BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF*, EPA-823-R00-012, Environmental Research Lab, Athens, GA.http://www.epa.gov/waterscience/basins/bsnsdocs.html#tech.

US Environmental Protection Agency (1998). Better Assessment Science Integrating Point and Nonpoint Sources (BASINS), User's Manual, USEPA, Environmental Research Lab, Washington, DC. http://www.epa.gov/waterscience/basins/bsnsdocs.html#buser.

US Environmental Protection Agency (1997).Guiding Principles for Monte Carlo Analysis, EPA/630/R-97/001. Offc. Of Water. US EPA: Washington, DC.

US Soil Survey Geographic Database (2006). National Resource Conservation Service (NRCS), Soil Data Mart, http://www.ncgc.nrcs.usda.gov/products/datasets/ssurgo/.

Van Liew, M.W., Arnold, J.G. and Garbrecht, J.D. (2003). Hydrological simulation on agricultural watersheds: Choosing between two models. *Trans. ASAE*, 46(6), 1539-1551.

Yakowitz, S., Krimmel, J.E. and Szidarovzky, F. (1978). Weighted Monte Carlo integration. *SIAM J. Numer. Anal.*, 15(6), 1289-1299.

Yin, Z., Walcoff, S., Kaplan, B., Cao, J., Lin, W., Chen, M., Liu, D. and Ning, Y. (2005). An analysis of the relationship between spatial patterns and water quality and urban development in Shanghai, China. *Comput., Environ. Urban Syst.,* 29, 197-221.

Yoder, D.C., Lyon, J.P. and Lightle, D.F. (2005). RUSLE2 Software NRCS Releases. http://fargo.nserl.purdue.edu/rusle2_dataweb/NRCS_Base_Database.htm.