A Stochastic Method to Characterize Model Uncertainty for a Nutrient TMDL

A. M. Sexton, A. Shirmohammadi, A. M. Sadeghi, H. J. Montas

Abstract. *The U.S. EPA's Total Maximum Daily Load (TMDL) program has encountered hindrances in its implementation partly because of its strong dependence on mathematical models to set limitations on the release of impairing substances. The uncertainty associated with predictions of such models is often not scientifically quantified and typically assigned as an arbitrary margin of safety (MOS) in the TMDL allocation. The Soil Water Assessment Tool (SWAT) model was evaluated to determine its applicability to identify the impairment status and tabulate a nutrient TMDL for a waterbody located in the Piedmont physiographic region of Maryland. The methodology for tabulating the nutrient TMDL is an enhancement over current methods used in Maryland. The mean-value first-order reliability method (MFORM) was paired with a stochastic approach to tabulate a science-based estimate of model uncertainty and MOS for the TMDL approach. Monthly streamflow estimates were quite good, with Nash-Sutcliffe efficiency (NSE) coefficients of 0.75 and 0.70 for the calibration and validation phases, respectively. Sediment and nutrients were not estimated as well as streamflow on a monthly basis; however, large improvements in model estimation were observed on an annual time scale. MOS was determined based on the desired level of confidence in meeting the water quality standard. The water quality standard was met at 20% nitrate reduction (9.9 kg N d^{-1}) with a 37.5% level of confidence. The water quality goal was met by a 30% reduction in nitrate load (8.6 kg N d^{-1}), in which case there was a 75% chance of meeting the water quality standard. Therefore, the MOS load (the difference between the standard and the goal) was 1.3 kg N d^{-1} or 10% of the baseline load. These results indicate that SWAT is a suitable model for use in TMDL assessments of impaired water bodies, especially assessments based on long-term simulations. In addition, the stochastic method used to quantify MOS for a nitrate TMDL is an improvement over current methods because it provides a formal, scientifically derived measure of model uncertainty.*

Keywords. *Calibration, Modeling, MFORM, MOS, SWAT, TMDL, Uncertainty.*

Mathematical models are mainly used in the Total Maximum Daily Load (TMDL) program to compensate for the scarcity and limitations of monitored data. Nonpoint-source pollution monitoring studies rarely have the ability to pinpoint sources of pollution and determine the best strategic plan to minimize pollution from different sources. Models, however, are able to estimate the amount of reduction necessary to meet water quality standards under different treatment scenarios as well as to predict the effect that treatment strategies have on water quality after implementation. One of the main problems associated with using mathematical models for TMDL assessment lies in the quantification of uncertainties (NRC, 2001; USEPA, 2002a). Stakeholders would like to have some sense of reliability in model predictions, especially when decisions based on model results can potentially impose both legal and financial responsibility upon point and nonpoint source contributors. Currently, the uncertainties in modeling are accounted for in the margin of safety (MOS) portion of TMDL allocations:

$$TMDL = \sum WLA + \sum LA + FG + MOS \qquad (1)$$

where *TMDL* represents the maximum amount of a pollutant that a waterbody can receive while still meeting its water quality standards, *WLA* represents waste load allocation for point sources (e.g., wastewater treatment plant discharges), *LA* corresponds to load allocation for nonpoint sources (e.g., agricultural runoff) and natural background contributions (e.g., weathering), *FG* represents future growth estimates of *WLA* and *LA*, and *MOS* accounts for uncertainty about pollutant loadings and waterbody response (USEPA, 1999a).

MOS is typically expressed in implicit or explicit terms (USEPA, 1999a). Implicit considerations involve making a conservative assumption when tabulating a TMDL, e.g., increasing the threshold of a water quality criterion above that which is necessary. Explicit considerations involve assigning a numeric safety factor to the value of MOS, e.g., 5% of the point and nonpoint source allocations. Both terms represent a highly subjective means of accounting for uncertainty (NRC, 2001).

Submitted for review in February 2011 as manuscript number SW 9054; approved for publication by the Soil & Water Division of ASABE in July 2011.

The authors are **Aisha M. Sexton, ASABE Member,** Visiting Scientist, Fischell Department of Bioengineering, University of Maryland, College Park, Maryland, and USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland; **Adel Shirmohammadi, ASABE Fellow,** Associate Dean and Professor, Department of Environmental Science and Technology and Fischell Department of Bioengineering, University of Maryland, College Park, Maryland; **Ali M. Sadeghi,** Soil Scientist, USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland; and **Hubert J. Montas, ASABE Member,** Associate Professor, Fischell Department of Bioengineering, University of Maryland, College Park, Maryland. **Corresponding author:** Aisha M. Sexton, U.S. EPA, 290 Broadway, New York, NY 10007; phone: 212-637-3608; e-mail: Aisha.Sexton@gmail.com.

EPA guidance and report documents (USEPA, 1999a, 1999b, 2002a) have suggested that MOS be calculated based on scientific information rather than subjectively assigned; however, only recently have scientists begun to devise and study formal uncertainty and error propagation strategies to determine MOS. In fact, after evaluating the current status of TMDL modeling technology, a multidisciplinary panel of experts proposed that the formal quantification of uncertainty be made an integral part of the TMDL process (Shirmohammadi et al., 2006). Hence, there is a need for further study and development of formal methods to calculate MOS.

Previous MOS Studies

Few researchers have attempted to translate the uncertainty in complex models used for TMDL analysis into MOS. Franceschini and Tsai (2008) used the modified Rosenblueth point estimate method to tabulate a reliability-based MOS. The method was applied to the Water Quality Analysis Simulation Program (WASP) model to calculate a TMDL for arsenic. Zhang and Yu (2004) applied first-order error analysis (FOEA) to another complex model, Hydrologic Simulation Program - Fortran (HSPF), to determine MOS for a nutrient TMDL. Both studies emphasized the use of uncertainty analysis methods requiring less computational time than the widely used Monte Carlo techniques. Computational efficiency becomes especially important if output uncertainty must be determined for multiple watersheds, as in the case of using reference watersheds (Wagner et al., 2007) or multiple land use change scenarios (Eckhardt et al., 2003) to develop TMDLs.

Walker (2003) developed a framework to tabulate MOS for lake phosphorus TMDLs by including stochastic terms in the phosphorus balance equation to reflect variability and uncertainty. The method is not easily applicable to complex watershed models, but the authors made some interesting observations about MOS tabulation. When both variability and uncertainty are considered, the result is often a large MOS value. This leads to higher costs for reduction measures and larger risk of not meeting water quality goals because of the uncertain performance level of many nonpoint-source control measures. Therefore, an adaptive implementation approach was suggested.

Borsuk et al. (2002) tabulated uncertainty of an empirical eutrophication model based on the exceedance frequency of a chlorophyll-*a* water quality standard. The confidence in meeting the percent-based standard and the amount of confidence required by watershed management decision makers were used to tabulate MOS. A similar approach was used in this study, although with application to a more complex mechanistic model.

A few of the MOS approaches discussed above are similar in that MOS is determined based on the level of confidence that the water quality standard will be met. This level of confidence can be a policy decision or determined by regulation, but it is a subjective quantity. Clearly, there must be some level of subjectivity in determining MOS and the extent to which MOS will be implemented, but the decisions should be based on proper scientific or deductive reasoning.

Use of SWAT in TMDL Studies

Previous studies using the Soil and Water Assessment Tool (SWAT) model (Arnold et al., 1998) have revealed a great deal of information about its performance and appropriate uses (Sexton et al., 2010; Tolson and Shoemaker, 2007; White and Chaubey, 2005; Chu and Shirmohammadi, 2004). However, further exploration is needed in the TMDL area, especially because the performance of TMDL modeling tools, including SWAT, seems to be inconsistent and in need of advancement (Borah et al., 2006; Muñoz-Carpena et al., 2006; Wagner et al., 2007). Such studies will demonstrate the applicability of SWAT to help identify impaired waterbodies, develop management scenarios to improve water quality conditions, and allocate restrictive loads to various contributing sources.

Santhi et al. (2001a) applied SWAT to the North Bosque River watershed in Texas to demonstrate the utility of the model in developing a TMDL for phosphorus. The model proved to be useful for representing the effects of alternative management scenarios for phosphorus reduction. Similar conclusions were also made for the Big Cypress Creek watershed in Texas (Santhi et al., 2003). Kang et al. (2006) used SWAT to develop TMDLs for suspended sediments, total nitrogen, and total phosphorus in a small watershed (385 ha) in Korea containing irrigated rice paddy fields. The total maximum daily load system (TOLOS) developed in the study was based on the integration of SWAT using geographic information system (GIS) and remote sensing (RS) data. The TOLOS was found to be a useful methodological tool for planning TMDLs.

Parajuli et al. (2008) used SWAT to demonstrate a targeted approach to effectively reduce fecal bacteria and sediment yield in the Upper Wakarusa watershed, a high-priority TMDL watershed in Kansas. Schilling and Wolter (2009) examined nitrate-nitrogen load reduction strategies using SWAT in the Des Moines River watershed. The required TMDL reduction of nitrate by 34.4% was met by reducing fertilizer application rate from 170 to 50 kg ha^{-1}. Targeted approaches were also examined and provided guidance as to the most efficient approaches for load reduction in the watershed. Radcliffe et al. (2009) demonstrated methods of estimating SWAT phosphorus (P) parameters for a lake north of Atlanta, Georgia, scheduled for a P TMDL. Models using P derived from the demonstrated methods outperformed models using default P parameter values.

As the literature indicates, a number of studies have tested SWAT for its usage in TMDL assessment. In addition, several studies have applied uncertainty analysis schemes to SWAT (Abbaspour et al., 2007; Arabi et al., 2007; Sohrabi, et al., 2003; Shen et al., 2008; Yang et al., 2008). However, our literature review found no studies that have demonstrated MOS calculation for TMDLs derived using SWAT.

The objective of this study was to present a scientifically based method of MOS tabulation using a formal and efficient uncertainty technique for use in SWAT and other complex water quality models used to establish TMDLs. The method was applied to a watershed in the Piedmont region of Maryland.

Materials and Methods

Site Description

Warner Creek watershed (~340 ha) is located in the Monocacy (fig. 1), a river basin that is known to contribute high levels of nutrients to the larger Chesapeake Bay watershed

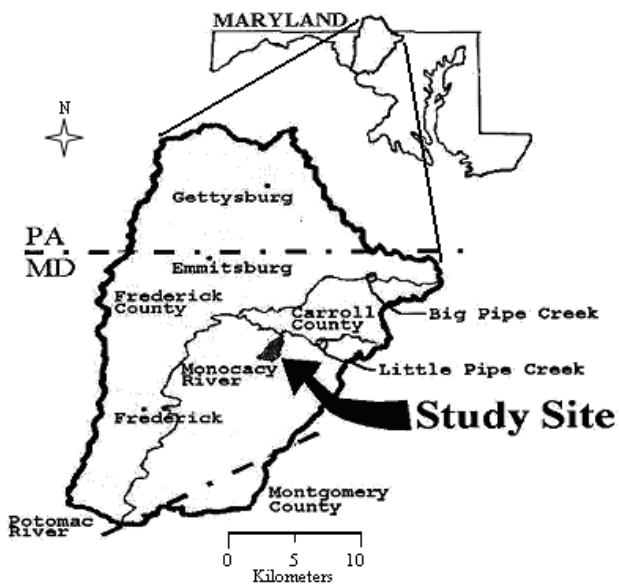

Figure 1. Location of Warner Creek watershed in Maryland.

(Blankenship, 2007; USDA-SCS, 1990). Several Monocacy water bodies were listed on Maryland's 303(d) list of impaired waters, including Lower Monocacy River, Upper Monocacy River, and Double Pipe Creek (MDE, 2008). Warner Creek is a tributary of Double Pipe Creek, which starts at the confluence of Big Pipe Creek and Little Pipe Creek (fig. 1). The majority of nutrient loads in the study watershed come from nonpoint sources, including loafing grounds for cattle and excess nutrients from croplands. Manor-Edgemont-Brandywine soils (~1/3 of watershed) and Penn-Reading-Croton soils (~2/3 of watershed) are the two dominant soil types in Warner Creek. All are well drained soils except for Croton, which is poorly drained. Most of the upland agricultural soils belong to the Penn silt loam series with slopes ranging from 3% to 8%. The land uses consist of mixed agriculture (~76%), urban (~13%), forest (~11%), and water (<1%).

To provide a general idea of climatic behavior at the study site, we examined annual precipitation amounts over the entire period of study (1994-2001), as shown in figure 2. The year 1996 was an unusually wet year with the state of Maryland receiving an average of 38% more rainfall than normal (USEPA, 2007). The northeastern blizzard of 1996, in which record snowfall occurred from January 6 to January 8, is a notable event (NOAA, 2011). Due to unusually warm weather and high flow volumes in mid-January (January 17 to January 19), major flooding occurred (fig. 3). Although annual yields do not reflect it, Maryland experienced drought conditions during the summer of 1998 through 1999 (Maryland State Climatologist Office, 2007). Another drought period began in May 2001 and lasted until December 2002.

Model Description and Data Acquisition

The SWAT water quality model (Arnold et al., 1998) is a watershed loading model that was developed by the USDA-ARS to estimate the impact of different management scenarios. It is a complex, physically based, semi-distributed model that operates in continuous time on a daily time step. The main components of the model include: climate, hydrology, land cover/plant growth, erosion, nutrients, pesticides, land management, channel and reservoir routing, soil temperature, and bacteria (Neitsch et al., 2005).

Daily rainfall and temperature data were obtained from measurements at the climate station set up by our group in subbasin 2 (fig. 4). Missing data were filled in using

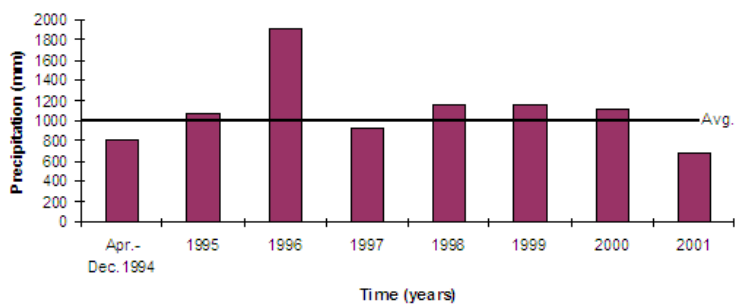

Figure 2. Annual precipitation and line of annual average precipitation in Warner Creek watershed.

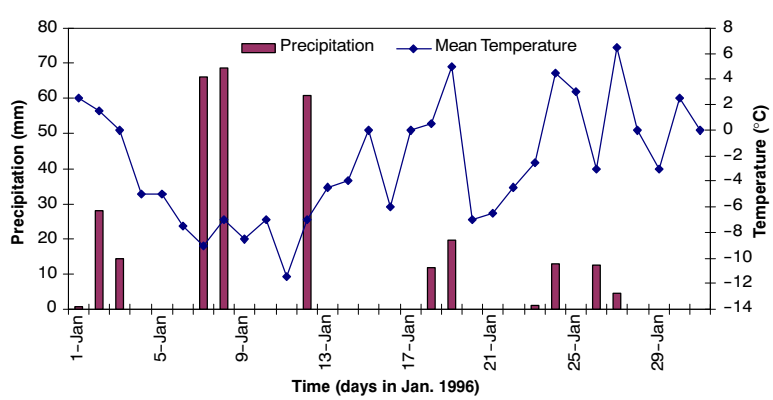

Figure 3. Warner Creek daily precipitation and mean temperatures in January 1996 (Maryland State Climatologist Office, 2007).

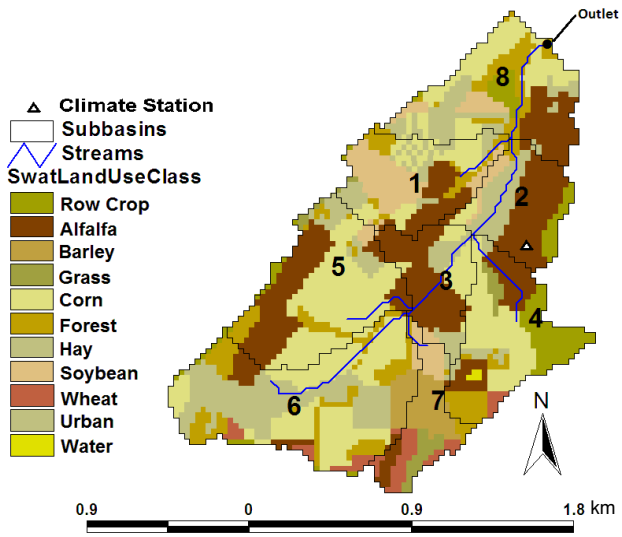

Figure 4. Location of subbasins and land use configuration.

measurements from a nearby monitoring station (~19 km away) in Emmitsburg, Maryland (fig. 1), obtained from the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC). Daily solar radiation, wind speed, and relative humidity data were generated using SWAT's weather generator. Observed flow and chemical loading data were obtained from field measurements; however, supplementary data were obtained from Chu and Shirmohammadi (2004) and Chu et al. (2004).

SWAT requires three GIS datasets in order to run: the digital elevation model (DEM), land cover/land use, and soils data. A 30 m resolution U.S. Geological Survey (USGS) National Elevation Dataset (NED) DEM was used. Land uses for each field were identified by aerial photos obtained from the USDA Agricultural Stabilization and Conservation Service (USDA-ASCS) office (Searing and Shirmohammadi, 1994). A Soil Survey Geographic (SSURGO) soils map of Frederick County, Maryland, was downloaded from the USDA-NRCS Soil Data Mart server. The watershed was delineated into eight subbasins based on the configuration of stream segments (fig. 4). Within each subbasin, the superimposing of similar land uses and soil types produced 53 hydrologic response units (HRUs).

Model Calibration and Validation

Model calibration and validation were conducted on streamflow, sediment, nitrate, and soluble phosphorus output. A baseflow filter program (Arnold et al., 1995; Arnold and Allen, 1999) was used to separate baseflow from measured streamflow in order to calibrate baseflow and surface runoff separately. Flow and nutrient loads were calibrated using approximately four years of data (April 1994-1997) and validated using another four years of data (1998-2001). Since sediment measurements were not available after 1997, sediment yield was calibrated using measured data from two years (April 1994 through 1995) and validated using 1996 and 1997 measured data. Manual calibration was performed by changing input parameters by percentage or absolute value. Input parameters used in calibration were chosen based on sensitivity analyses found in the literature (Chu and Shirmohammadi, 2004; Chu et al., 2004; Sohrabi et al., 2003; White and Chaubey, 2005) and their physical relation to output variables.

Model Performance Criteria

Moriasi et al. (2007) introduced a set of recommended criteria to establish systematic and universal guidelines for model evaluation. The recommended criteria and some additional evaluation criteria were used in this study, including square of the correlation coefficient (r^2), Nash-Sutcliffe efficiency coefficient (NSE), ratio of root mean square error to observations' standard deviation (RSR), and percent bias (PBIAS):

$$r^2 = \left[\frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^{n} (P_i - \bar{P})^2}} \right]^2 \quad (2)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2} \quad (3)$$

$$\mathrm{RSR} = \frac{\mathrm{RMSE}}{\mathrm{SD}_{O_i}} = \frac{\sqrt{\sum_{i=1}^{n} (O_i - P_i)^2}}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2}} \quad (4)$$

$$\mathrm{PBIAS} = \left[\frac{\sum_{i=1}^{n} (O_i - P_i) * 100}{\sum_{i=1}^{n} O_i} \right] (\%) \quad (5)$$

Table 1. Description of SWAT input parameters (basic variables) evaluated in uncertainty analysis of model nitrate estimation.

Parameter	Description
ANION_EXCL	Fraction of porosity from which anions are excluded
BIOMIX	Biological mixing efficiency
CMN	Rate factor for humus mineralization of active organic nutrients (N and P)
FRT_SURF	Fraction of fertilizer applied to the top 10 mm of soil
SOL_NO3_1	Initial NO_3 concentration in soil layer 1 (mg N kg^{-1})
SOL_NO3_2	Initial NO_3 concentration in soil layer 2 (mg N kg^{-1})
NPERCO	Nitrate percolation coefficient

where O_i are observed and P_i are predicted data, $\overline{O}$ and $\overline{P}$ are observed and predicted mean values, SD is standard deviation, and n is the number of samples.

Uncertainty Analysis Method

The mean-value first-order reliability method (MFORM; Melching and Bauwens, 2001) was chosen to quantify uncertainties in the model prediction of nitrate concentration after model calibration. This approach allows the user to determine the variance in the output variable as well as the variance contributed by each important input parameter, otherwise known as basic variables. Basic variables were determined to be important based on sensitivity analyses found in the literature (Sohrabi et al., 2003; White and Chaubey, 2005), the physical meaning of variables as they relate to output variables, and the level of parameter importance during model calibration. A description of each basic variable considered in the uncertainty of nitrate is listed in table 1. Uncertainties quantified for streamflow, sediment and phosphate outputs along with their basic variables were recorded by Sexton et al. (2011).

MFORM is derived by performing a Taylor series expansion of the model output function as follows:

$$Y = g(X_e) + \sum_{i=1}^{n} (x_i - x_{ie}) \left(\frac{\partial g}{\partial x_i} \right)_{X_e} \quad (6)$$

where Y is the dependent variable or model output of interest, $g(X_e)$ is the function representing the simulation process to obtain Y, X_e is the vector of basic variables at the expansion point, n is the number of basic variables x_i, and $\partial g / \partial x_i$ represents the rate of change of the model output with respect to a unit change in each basic variable, usually referred to as the sensitivity coefficient. In MFORM, the expansion point is at the mean value of basic variables. Therefore, the mean and variance of the dependent variable can be approximated as:

$$E(Y) \approx g(X_m) \quad (7)$$

$$Var(Y) = \sigma_Y{}^2 \approx \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i} \right)_{X_m}^2 \sigma_i{}^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial g}{\partial x_i} \right)_{X_m} \left(\frac{\partial g}{\partial x_j} \right)_{X_m} \cdot C_v(x_i, x_j) \quad (8)$$

where $E(Y)$ is the expected value (mean) of random variable Y, X_m is the vector of basic variables at the mean values, $\sigma_i{}^2$ is the variance of basic variable i, $C_v(x_i, x_j)$ is the covariance of basic variables i and j, and all other variables are as previously defined. The first term in equation 8 represents the variance of statistically independent parameters, while the second term is used to tabulate the variance of correlated parameters. In many cases, the relationship between parameters is unknown. Therefore, the parameters are often assumed to have no correlation (Melching and Yoon, 1996; Zhang and Yu, 2004; Sexton et al, 2011). However, neglecting parameter correlations could result in an underestimation of parameter and output uncertainty depending on the significance of correlations. MOS and confidence of compliance would, in turn, possibly be underestimated as well. If basic variables are not correlated, then $C_v(x_i, x_j)$ is equal to zero. In this case, the variance of output can be written as:

$$Var(Y) = \sigma_Y{}^2 \approx \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i} \right)_{X_m}^2 \sigma_i{}^2 \quad (9)$$

This term represents the fraction of model output variance (FOV) contributed by each basic variable (x_i). In this equation the squared sensitivity coefficient ($\partial g / \partial x_i$) serves as a way to assign a measure of importance to the variance of each basic variable. When using complex models, the best way to solve for $\partial g / \partial x_i$ is by using numerical methods. In this study, a forward difference numerical scheme was used in which each parameter was increased by 5%. The output of MFORM, in this case, was the variance of daily nitrate concentrations simulated in SWAT. Daily variances were converted to standard deviations (square root of the variance) to be utilized in the stochastic methodology to tabulate MOS, as described in the next section.

Margin of Safety (MOS) and TMDL Estimation

EPA guidance has suggested making probability-based water quality impairment decisions for conventional pollutants (USEPA, 1997). The purpose of such guidance was to account for measurement error and potentially small data sets not properly representing the conditions of a waterbody (USEPA, 2003a). EPA guidance and recommendations from other agencies and scientists have also stressed the importance of using formal uncertainty analysis methods to tabulate MOS as opposed to arbitrarily assigning its value (NRC, 2001; Shirmohammadi et al., 2006; USEPA, 2002a).

Current methods of nutrient TMDL assessment in Maryland are not probability-based and do not account for MOS using a formal uncertainty analysis scheme. MOS is determined explicitly and is typically assigned as 5% to 10% of nonpoint-source load allocations. This study suggests an alternative method of nutrient TMDL assessment, which uses a probability-based or stochastic approach along with MFORM to determine MOS and TMDL.

The designated use of the study waterbody is aquatic life support (Shirmohammadi and Montas, 2003). That is interpreted here as Maryland's Use I designation (Code of Maryland Regulations [COMAR] 26.08.02.02), which is water-contact recreation, and protection of non-tidal warm-water aquatic life. In Maryland, a waterbody of this use would normally be evaluated for nutrient contamination using indicators such as chlorophyll-*a* and dissolved oxygen concentration; however, data containing that information were not collected in Warner Creek watershed. We therefore assumed a Use I-P designation, which is the same as Use I

mentioned above but contains an additional use as a public water supply, hence the P classification. This justified the employment of nitrate concentration (a drinking water contaminant) as an indicator of waterbody impairment in this study.

According to EPA's national recommended water quality criteria, the water quality criterion for nitrate concentration measured as nitrogen is 10 mg L^{-1} for a waterbody designated for drinking water use (USEPA, 2003b). The same criterion is used in Maryland. EPA guidance has recommended listing a waterbody as impaired if greater than 10% of conventional chemical samples exceed the assigned water quality criterion (USEPA, 1997, 2002b). This type of probability-based standard is meant to account for natural variability and measurement error (Borsuk et al., 2002; USEPA, 2003a). Maryland determines that a waterbody is impaired by chemical contaminants if greater than 10% of samples exceed the criteria, with a minimum of ten samples over a three-year period (MDE, 2010). In this study, we used daily samples from 1994 to 2001.

After examining the number of daily nitrate concentrations that exceeded the maximum contaminant level (MCL) of 10 mg L^{-1} in Warner Creek watershed, we determined that the waterbody was not impaired. Less than 10% of daily nitrate samples exceeded the 10 mg L^{-1} MCL over the entire time period (1994-2001); therefore, EPA guidelines were met. Best management practices (BMPs) implemented in this watershed prior to sampling efforts may be attributed to the unimpaired status of the waterbody. To create an impaired waterbody scenario, we assumed a nitrate MCL of 6 mg L^{-1}. This increased the probability of exceedance so the MOS/TMDL methodology could be properly demonstrated.

Daily nitrate concentrations computed by the calibrated SWAT model (baseline conditions representing current conditions of the waterbody) and the associated daily standard deviations tabulated with MFORM were used to calculate daily exceedance probabilities as (Borsuk et al., 2002):

$$p = P(c > c^* \mid \beta, \sigma, X) = 1 - F\left(\frac{c^* - g(X, \beta)}{\sigma}\right) \quad (10)$$

where p is the exceedance probability, c is the chemical concentration, c^* is the numerical criterion of c (6 mg L^{-1}), β represents the set of all model parameters (e.g., curve number and hydraulic conductivity), σ is the standard deviation of chemical concentration (found using MFORM), X represents model input variables (e.g., precipitation and temperature), $g(X, \beta)$ is the output chemical concentration generated by the model (also known as c), and the function F depicts the cumulative standard normal distribution.

Days within the wet season were considered the critical period in each year. The likelihood of exceeding the water quality standard was larger during that time frame, which represented the worst-case scenario. The critical period of each year was then used to tabulate the exceedance frequency. Exceedance frequency is defined as the number of days that the exceedance probability is greater than 10% divided by the total number of days in the critical time period of each year. A probability distribution of annual exceedance frequencies was formulated to describe the uncertainty in the exceedance frequency resulting from parameter uncertainty. The expected exceedance is defined as the mean of the distribution of annual frequency values.

The portion of the probability distribution of exceedance frequencies less than or equal to 10% represents the probability that the true exceedance frequency will meet the 10% frequency standard. Borsuk et al. (2002) referred to that portion of the distribution as the confidence of compliance (CC). CC is a measure by which water quality goals can be expressed. For example, if a water quality manager wanted to be 40% confident that the exceedance frequency in a waterbody is 10% or less, the manager would then reduce the load until the CC goal of 40% was met.

Once CC was determined for the baseline nitrate load (no reductions), several other CCs were obtained by reducing the percentage of load flowing into the waterbody (5% up to 40%). The nitrate load associated with the desired CC to meet the water quality goal was then compared to the load required to meet the water quality standard. The difference between the load required to meet the water quality standard and the load required to meet the water quality goal was assigned to the margin of safety value. The TMDL for the waterbody is considered to be the resultant load after the load has been reduced enough to meet the water quality standard.

Results and Discussion

Hydrology

Monthly streamflow estimation was very good during the calibration period, with resulting NSE, RSR, and PBIAS values of 0.75, 0.49, and -15.1%, respectively (table 2). Model performance during streamflow validation was slightly lower than during the calibration period but could still be categorized as good, with NSE and RSR values of 0.70 and 0.54, respectively. Although streamflow during the validation period was overpredicted for a major portion of the test period, model performance considering systematic deviations was very good because PBIAS was within the range of ±10%. Total streamflow measurements are a combination of surface and baseflow values. Model performance was clearly worst during baseflow calibration compared to surface runoff calibration (table 2). One reason for this is because errors associated with high flow values tend to be larger than those associated with low flow values, especially when squared terms, e.g., the term $(O_i - \overline{O})^2$, are used in optimization or evaluation criteria, such as r^2 (eq. 2) and NSE (eq. 3). Therefore, trying to minimize high flow errors often leads to fitting the higher portion of the hydrograph (i.e., peak surface flows) at the expense of lower portions (i.e., baseflow) (Krause et al., 2005). In addition, inconsistencies in flow prediction may be attributed to the use of empirical information to derive flow using the SCS curve number method and errors in sample measurement and transcription.

Sediment

Monthly sediment yield estimations during the calibration and validation periods were poor, with NSE values around 0.2 (table 3). However, an NSE value greater than zero means that the model is a better predictor of measured data than using the mean of observed values. Results for monthly sediment prediction were unsatisfactory due to a number of reasons. Measured data were lacking due to sparse measurements, unforeseen occurrences in watershed management (e.g., cows wandering through the stream), and high flow volumes resulting from above-normal temperatures (see fig. 3)

Table 2. Summary statistics and model evaluation results for monthly hydrology during the calibration and validation periods.[a]

Monthly Hydrologic Measurement			Mean (mm)	SD (mm)	No. of Samples	r^2	b	RMSE (mm)	NSE	RSR	PBIAS (%)
Calibration Period (April 1994-1997)	Surface runoff	Measured	19.74	37.07	45	0.73	0.38	24.62	0.55	0.66	18.50
		Simulated	16.09	16.24	45						
	Baseflow	Measured	19.50	17.19	45	0.30	0.92	26.23	-1.38	1.53	-49.20
		Simulated	29.10	29.38	45						
	Streamflow	Measured	39.24	46.53	45	0.79	0.87	22.82	0.75	0.49	-15.10
		Simulated	45.18	45.62	45						
Validation Period (1998-2001)	Surface runoff	Measured	11.30	18.45	48	0.68	0.60	10.57	0.67	0.57	-5.50
		Simulated	11.92	13.37	48						
	Baseflow	Measured	19.34	24.79	48	0.49	0.69	18.80	0.41	0.76	-11.50
		Simulated	21.56	24.18	48						
	Streamflow	Measured	30.58	38.25	48	0.73	0.84	20.64	0.70	0.54	-9.50
		Simulated	33.48	37.54	48						

[a] SD = standard deviation, r^2 = coefficient of determination, b = slope, RMSE = root mean square error, NSE = Nash-Sutcliffe coefficient of efficiency, RSR = RMSE-observation's standard deviation ratio, and PBIAS = percent bias.

Table 3. Summary statistics and model evaluation criteria results for monthly sediment loadings during the calibration and validation periods, and annual loads over the entire time period.[a]

Sediment Measurement		Mean (kg ha^{-1})	SD (kg ha^{-1})	No. of Samples	r^2	b	RMSE (kg ha^{-1})	NSE	RSR	PBIAS (%)
Calibration period (April 1994-1995), Monthly sediment yield	Measured	324.40	850.38	21	0.47	0.14	743.33	0.20	0.87	48.5
	Simulated	167.21	167.92	21						
Validation period (1996-1997), Monthly sediment yield	Measured	244.38	434.30	24	0.24	0.3	379.15	0.21	0.87	-22.3
	Simulated	298.96	266.31	24						
Entire period (April 1994-1997), Annual sediment yield	Measured	3169.36	2409.34	4	0.63	0.71	1374.59	0.57	0.57	15.7
	Simulated	2671.58	2139.43	4						

[a] SD = standard deviation, r^2 = coefficient of determination, b = slope, RMSE = root mean square error, NSE = Nash-Sutcliffe coefficient of efficiency, RSR = RMSE-observation's standard deviation ratio, and PBIAS = percent bias.

after snowfall (Maryland State Climatologist Office, 2007). In addition, daily samples were aggregated to monthly averages, in which case there is a loss of information and increase in absolute errors, resulting in measurements further away from the true load (Harmel and King, 2005). Trends during 1996 were unsatisfactory, mainly because 1996 was an extremely wet year (fig. 2). The SWAT model does not seem to perform well in predicting peak flow and sediment yield under extremely wet conditions or extreme storm events (Borah and Bera, 2004). Some of the above-mentioned issues have been attributed to the use of empirical equations and insufficient routing mechanisms in the model (Sexton et al., 2011; Borah et al., 2006). Similarly, poor predictions of monthly sediment yield using SWAT have been reported in the literature (Kirsch et al., 2002; Santhi et al., 2001b).

Based on the RSR, PBIAS, and NSE values of 0.57, 15.7%, and 0.57, respectively (table 3), the annual sediment yield results can be considered satisfactory. Annual predictions were expected to be an improvement over monthly predictions since SWAT was designed for long-term simulation and assessment of the relative impact of BMPs and not for short-term events.

Nutrients

Nitrate calibration trends were fairly good, with the exception of extremely wet conditions in 1996. A high nitrate loading event in January 1996 was attributed to unusually high flow events as well as ammonium-based deicer placed on county roads during that time frame. The ammonium-based deicers eventually oxidize to nitrate (B. Dixon, 2001, personal communication on ammonium-based deicers, College Park, Maryland). Nitrate load was underestimated in areas where streamflow was also underestimated, pointing to the close association between nitrate discharge and flow. This was true during the validation period as well. The NSE and RSR performance criteria can both be interpreted as unsatisfactory, with values of 0.35 and 0.80, respectively, during calibration and 0.44 and 0.74, respectively, during validation (table 4).

Discrepancies in baseflow measurement could also have contributed to poor nitrate prediction. Measured nutrient loadings were adjusted to remove the chemical contribution transported by subsurface flow from outside of the watershed (Chu and Shirmohammadi, 2004). Similar performance of the SWAT model for monthly nitrate prediction was observed in the literature (Saleh et al., 2000; Santhi et al., 2001b). Annual results of nitrate loading over the entire eight-year period show good model performance based on NSE and RSR values of 0.67 and 0.54, respectively. A PBIAS of +20.6 for annual results represents relatively good model performance in terms of the average tendency of the simulated data to be larger or smaller than the observed data.

Soluble phosphorous (sol P) was predicted very poorly during the calibration period. Sol P in January 1996 was largely overpredicted. Later, in November 1997, an abnormally high sol P load was observed. Those observations were likely due to measurement errors. As mentioned earlier, measured sediment data were very poor in quality, and since the organic portion of phosphorous mainly travels attached to sediment, this may have lead to inconsistencies in sol P mea–

Table 4. Summary statistics and model evaluation criteria results for monthly nitrate loadings during the calibration and validation periods, and annual loads over the entire time period.[a]

Nitrate Measurement		Mean (kg ha^{-1})	SD (kg ha^{-1})	No. of Samples	r^2	b	RMSE (kg ha^{-1})	NSE	RSR	PBIAS (%)
Calibration period (April 1994-1997), Monthly nitrate load	Measured	1.58	1.77	45	0.40	0.41	1.41	0.35	0.80	25.7
	Simulated	1.18	1.14	45						
Validation period (1998-2001), Monthly nitrate load	Measured	1.21	1.53	48	0.50	0.66	1.13	0.44	0.74	14.2
	Simulated	1.03	1.42	48						
Entire period (April 1994-2001), Annual nitrate load	Measured	16.14	7.97	8	0.88	0.80	4.27	0.67	0.54	20.6
	Simulated	12.82	6.80	8						

[a] SD = standard deviation, r^2 = coefficient of determination, b = slope, RMSE = root mean square error, NSE = Nash-Sutcliffe coefficient of efficiency, RSR = RMSE-observation's standard deviation ratio, and PBIAS = percent bias.

Table 5. Summary statistics and model evaluation criteria results for monthly phosphate loadings during the calibration and validation periods, and annual loads over the entire time period.[a]

Phosphate Measurement		Mean (kg ha^{-1})	SD (kg ha^{-1})	No. of Samples	r^2	b	RMSE (kg ha^{-1})	NSE	RSR	PBIAS (%)
Calibration period (April 1994-1997), Monthly phosphate load	Measured	0.34	0.51	45	0.14	0.42	0.61	-0.47	1.20	0.0
	Simulated	0.34	0.59	45						
Validation period (1998-2001), Monthly phosphate load	Measured	0.23	0.35	48	0.59	0.91	0.27	0.41	0.76	-8.2
	Simulated	0.25	0.42	48						
Entire period (April 1994-2001), Annual phosphate load	Measured	3.34	1.64	8	0.64	1.30	1.58	-0.06	0.96	-3.5
	Simulated	3.46	2.67	8						

[a] SD = standard deviation, r2 = coefficient of determination, b = slope, RMSE = root mean square error, NSE = Nash-Sutcliffe coefficient of efficiency, RSR = RMSE-observation's standard deviation ratio, and PBIAS = percent bias.

surement. For instance, neither sediment nor phosphorus measurements reflected the high flow volumes that resulted days after the blizzard of 1996. Therefore, the model overpredicted those measurements.

Monthly predictions during the validation period were much better than during calibration. Model evaluation criteria reflect this difference, with NSE and RSR values of -0.47 and 1.2, respectively, during the calibration period and 0.41 and 0.76, respectively, during the validation period (table 5). This is most likely due to the poor model performance during 1996, the wettest year that was simulated in the calibration period. If 1996 is removed from the calibration period, the NSE (0.18) and RSR (089) values are improved considerably. Although the NSE and RSR values during validation on a monthly basis are unsatisfactory, the r^2 value of 0.59 indicates moderate model performance. The slope of 0.91 (table 5) also represents a positive rating. Similar performance of the SWAT model for monthly sol P prediction was observed in the literature (Chu et al., 2004). Annual results over the eight- year period were an improvement over calibrated monthly results but not over the validated monthly results (table 5).

MFORM Uncertainty Analysis

Uncertainty outputs (recorded by Sexton et al., 2011) upheld the above model performance results. The largest amount of variance in output variables occurred during wet periods. Predicted sediment output had the largest amount of variability around its mean, followed by nitrate, phosphate, and streamflow as indicated by average annual coefficients of variation of 28%, 19%, 17%, and 15%, respectively. Of the parameters listed in table 1, ANION_EXCL contributed the largest amount of total variance (96%) to nitrate output. BIOMIX contributed about 3% of the variance, and the remaining parameter contributions were minute (Sexton et al., 2011). As mentioned in the Materials ands Methods section, daily variances of nitrate concentration (derived using MFORM) were plugged into equation 10 to calculate exceedance probability and further tabulate the MOS value for the TMDL.

TMDL and Margin of Safety

Table 6 shows the average daily model predictions of nitrate load and concentration, the expected exceedance frequency, and the confidence of compliance for each load reduction of nitrate from 0% to 40% over the entire study period (1994-2001). Nitrate load reductions greater than 40% resulted in a confidence of compliance equal to 100%; therefore, load reductions greater than 40% are not listed. The expected exceedance of 14% at 0% load reduction indicated impaired waterbody status. At 0% nitrate load reduction, there was only 12.5% confidence that the 10% frequency standard would be met. At a load reduction of 20%, the mean exceedance frequency was expected to be 10% with a confidence of compliance equal to 37.5%, which was the load re–

Table 6. Average daily model predictions of nitrate load and concentration, expected exceedance frequency, and confidence of compliance for each load reduction of nitrate from 0% to 40% over the entire study period (1994-2001).

NO_3-N Reduction (%)	Average NO_3-N Load (kg N d^{-1})	Average NO_3-N Conc. (mg L^{-1})	Expected Exceedances (%)	Confidence of Compliance (%)
0	12.3	1.25	14	12.5
5	11.7	1.19	12	25
10	11.1	1.12	11	25
15	10.5	1.06	11	37.5
20	9.9	1.00	10	37.5
25	9.3	0.94	9	37.5
30	8.6	0.87	8	75
35	8.0	0.81	8	75
40	7.4	0.75	7	100

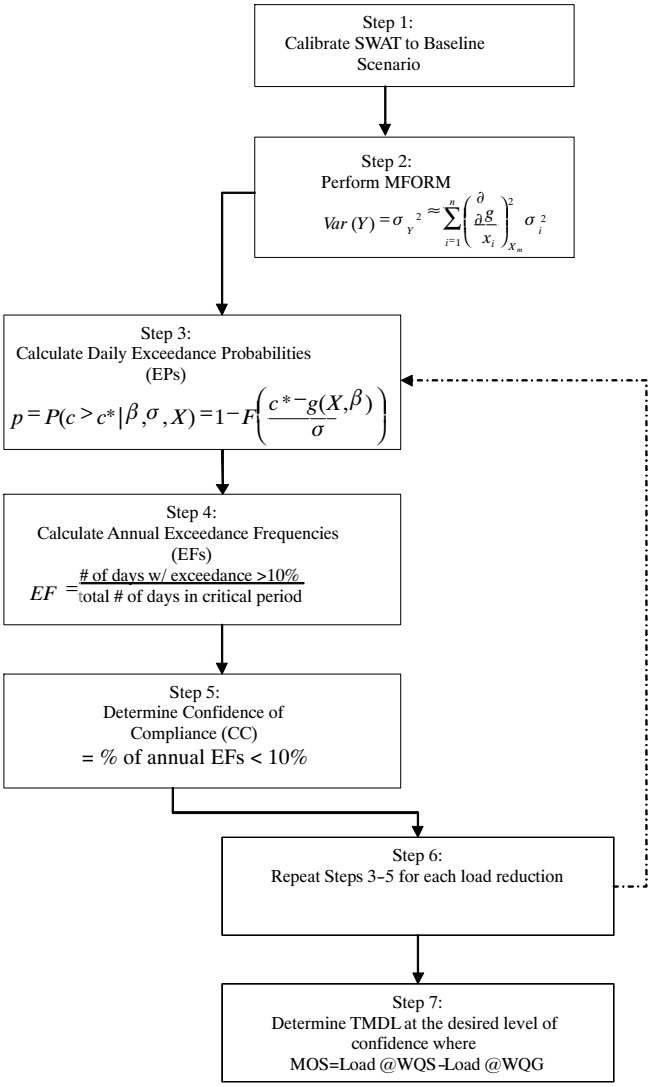

Figure 5. Flowchart of steps for using the SWAT model and uncertainty analysis in TMDL assessment.

duction necessary to meet the nitrate water quality standard. MOS was then determined based on the desired level of confidence. Therefore, at a desired level of confidence of 75%, the MOS load was equal to the difference between the load at 20% reduction (9.9 kg N d^{-1}) and the load at 30% reduction (8.6 kg N d^{-1}), which was 1.3 kg N d^{-1}. This reduced the first portion of equation 1 ($\Sigma\ WLA + \Sigma\ LA + FG$) to 8.6 kg N d^{-1}. The TMDL for this waterbody was therefore 9.9 kg N d^{-1} (TMDL= 8.6 kg N d^{-1} + 1.3 kg N d^{-1}).

The probability that the true exceedance frequency will be below 10% is 100% at the nitrate reduction level of 40%. Using a water quality criterion of 6 mg L^{-1} did not make the water body at the outlet of the Warner Creek watershed extremely impaired. However, in other cases when a water body is highly impaired, it may not be as feasible to set a water quality goal to 75% confidence, especially when the effectiveness and efficiency of improvement strategies are not well known. In that case, many researchers have suggested an adaptive management approach in which an initial set of improvement strategies are implemented, monitored, and evaluated to determine their efficacy, and then further measures can be taken (Dilks and Freedman, 2004; Walker, 2003).

In summary, this study suggests the following steps for using the SWAT model for TMDL assessment. These steps are presented as a flowchart depicted in figure 5:

1. Calibrate the model to represent observed conditions of the impaired waterbody. In the absence of measured data for the watershed of interest, data from a reference watershed sharing the same physiographic region, land use, and climatic conditions may be used to calibrate the model. This is the baseline scenario.
2. Perform MFORM methodology using mean values and standard deviations of input parameters to determine daily average standard deviations of the output variable concentrations.
3. Use daily average standard deviations obtained from the MFORM method in step 2, and calibrated SWAT's output concentrations to calculate daily exceedance probabilities. Equation 10 represents this step.
4. Calculate the annual exceedance frequencies (EFs) (number of days with exceedances over 10% divided by the total number of days in the critical period).
5. Determine the percent of annual EFs that are less than or equal to 10% (confidence of compliance).
6. Repeat steps 3 through 5 for each incremental reduction of nutrient load until the confidence of compliance reaches 100%. It should be noted that different reduction levels may be needed for different watersheds based on their level of impairment.
7. Evaluate each level of confidence with its corresponding nutrient reduction to set a feasible water quality goal (WQG). The difference in load between the nutrient reduction that meets water quality standards (WQS) (an expected exceedance less than or equal to 10%) and the nutrient reduction that meets the water quality goal (desired level of confidence) should be assigned to the MOS value.

Summary and Conclusions

In this study, SWAT was a good predictor of total streamflow, a poor predictor of sediment yield, and at best a moderate predictor of nutrient loadings on a monthly basis. Model performance measures were upheld by uncertainty analysis, with sediment output having the largest average annual coefficient of variation and streamflow prediction having the smallest coefficient of variation. More knowledge about the true values of input parameters and measured data will lead to better model performance and less output uncertainty. More attention is also needed on the model routing mechanisms and structure in order to more properly handle environmental processes. Performance levels improved for annual output results, demonstrating the most appropriate use of the model for long-term simulation studies.

The methodology used in this study to determine MOS and TMDL for a nitrate-contaminated waterbody included MFORM paired with a stochastic approach. Initial SWAT results indicated that the waterbody was not impaired. Therefore, a hypothetical TMDL scenario was set up (by assuming a lower MCL of 6 mg L^{-1}) in order to properly demonstrate the method. Results produced an MOS of 1.3 kg N d^{-1} and a TMDL equal to 9.9 kg N d^{-1}. This method is an improvement over current methods employed in Maryland and other states because it is probability-based and includes a formal method

of uncertainty analysis, which gives stakeholders an idea of the reliability of the model results.

Although a confidence of compliance of 100% was reached at a nitrate load reduction of 40% for Warner Creek watershed, this level of confidence may not be as feasible for highly impaired waterbodies. In that case, an adaptive management strategy could prove to be more effective by identifying and implementing the most useful management practices in a stepwise process, as opposed to spending a lot of money to implement a large slate of BMPs to no effect. Future suggested work is to apply this method to an impaired waterbody listed on Maryland's 303(d) list to demonstrate its impaired status using SWAT along with its TMDL solution using the method of this study with regard to actual Maryland regulations. The feasibility of confidence levels can be examined in this regard as well.

Acknowledgements

The authors would like to thank Jeff Arnold, Nancy Sammons, Mauro Di Luzio, and Daniel Moriasi at Blackland Research Center in Temple, Texas, for their advice and support during model configuration and calibration.

References

Abbaspour, K. C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan. 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *J. Hydrol.* 333(2-4): 413-430.

Arabi, M., R. S. Govindaraju, and M. M. Hantush. 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. *J. Hydrol.* 333(2-4): 459-471.

Arnold, J. G., and P. M. Allen. 1999. Automated methods for estimating baseflow and groundwater recharge from streamflow records. *J. American Water Resources Assoc.* 35(2): 411-424.

Arnold, J. G., P. M. Allen, R. S. Muttiah, and G. Bernhardt. 1995. Automated baseflow separation and recession analysis techniques. *Groundwater* 33(6): 1010-1018.

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

Blankenship, K. 2007. Out of shape: No matter how one looks at it, bay's health fails to make the grade. *Chesapeake Bay J.* 17(3): 1, 10-12.

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

Borah, D. K., G. Yagow, A. Saleh, P. L. Barnes, W. Rosenthal, E. C. Krug, and L. M. Hauck. 2006. Sediment and nutrient modeling for TMDL development and implementation. *Trans. ASABE* 49(4): 967-986.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. *Environ. Sci. and Tech.* 36(10): 2109-2115.

Chu, T. W., and A. Shirmohammadi. 2004. Evaluation of the SWAT model's hydrology component in the Piedmont physiographic region of Maryland. *Trans. ASAE* 47(4): 1057-1073.

Chu, T. W., A. Shirmohammadi, H. Montas, and A. Sadeghi. 2004. Evaluation of the SWAT model's sediment and nutrient components in the Piedmont physiographic region of Maryland. *Trans. ASAE* 47(5): 1523-1538.

Dilks, D. W., and P. L. Freedman. 2004. Improved consideration of the margin of safety in total maximum daily load development. *J. Environ. Eng.* 130(6): 690-694.

Eckhardt, K., L. Breuer, and H.-G. Frede. 2003. Parameter uncertainty and the significance of simulated land use change effects. *J. Hydrol.* 273: 164-176.

Franceschini, S., and C. W. Tsai. 2008. Incorporating reliability into the definition of the margin of safety in total maximum daily load calculations. *J. Water Resources Planning and Mgmt.* 134(1): 34-44.

Harmel, R. D., and K. W. King. 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. *Trans. ASABE* 48(5): 1713-1721.

Kang, M. S., S. W. Park, J. J. Lee, and K. H. Yoo. 2006. Applying SWAT for TMDL programs to a small watershed containing rice paddy fields. *Agric. Water Mgmt.* 79(1): 72-92.

Kirsch, K. J., A. Kirsch, and J. G. Arnold. 2002. Predicting sediment and phosphorus loads in the Rock River basin using SWAT. *Trans. ASAE* 45(6): 1757-1769.

Krause, P., D. P. Boyle, and F. Base. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosci.* 5: 89-97.

Maryland State Climatologist Office. 2007. 1998-1999 drought information. College Park, Md.: University of Maryland, Maryland State Climatologist Office.

MDE. 2008. The 2008 integrated report of surface water quality in Maryland. Baltimore, Md.: Maryland Department of the Environment. Available at: www.mde.state.md.us/programs/Water/TMDL/Integrated303dReports/Documents/wwwmde.state.md.us/assets/document/2008_IR_Parts_A_thru_E(1).pdf. Accessed 3 March 2011.

MDE. 2010. Methodology for determining impaired waters by chemical contaminants for Maryland's integrated report of surface water quality. Baltimore, Md.: Maryland Department of the Environment. Available at: www.mde.state.md.us/programs/Water/TMDL/Integrated303dReports/Documents/wwwmde.state.md.us/assets/document/Toxics_Listing_Methodology_2010.pdf. Accessed 16 May 2011.

Melching, C. S., and W. Bauwens. 2001. Uncertainty in coupled nonpoint source and stream water-quality models. *J. Water Resources Planning and Mgmt.* 127(6): 403-413.

Melching, C. S., and C. G. Yoon. 1996. Key source of uncertainty in QUAL2E model of Passaic River. *J. Water Resources Planning and Mgmt.* 122(2): 105-113.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50(3): 885-900.

Muñoz-Carpena, R., G. Vellidis, A. Shirmohammadi, and W. W. Wallender. 2006. Evaluation of modeling tools for TMDL development and implementation. *Trans. ASABE* 49(4): 961-965.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams. 2005. *Soil and Water Assessment Tool Theoretical Documentation.* Version 2005. Temple, Tex.: USDA-ARS Grassland, Soil and Water Research Laboratory and Blackland Research Center.

NOAA. 2011. Northeast floods of 1996 disaster survey report table of contents. Silver Spring, Md.: National Oceanic and Atmospheric Administration, National Weather Service. Available at: www.weather.gov/oh/hrl/surveys/flood96/contents.htm. Accessed 4 May 2011.

NRC. 2001. Assessing the TMDL approach to water quality management. Washington, D.C.: National Research Council.

Parajuli, P. B., K. R. Mankin, and P. L. Barnes. 2008. Applicability of targeting vegetative filter strips to abate fecal bacteria and sediment yield using SWAT. *Agric. Water Mgmt.* 95(10): 1189-1200.

Radcliffe, D. E., Z. Lin, L. M. Risse, J. J. Romeis, and C. R. Jackson. 2009. Modeling phosphorus in the Lake Allatoona

watershed using SWAT: I. Developing phosphorus parameter values. *J. Environ. Qual.* 38(1): 111-120.

Saleh, A., J. G. Arnold, P. W. Gassman, L. M. Hauk, W. D. Rosenthal, J. R. Williams, and A. M. S. MacFarland. 2000. Application of SWAT for the upper North Bosque River watershed. *Trans. ASABE* 43(5): 1077-1087.

Santhi, C., J. G. Arnold, J. R. Williams, L. M. Hauck, and W. A. Gugas. 2001a. Application of a watershed model to evaluate management effects on point and nonpoint source pollution. *Trans. ASAE* 44(6): 1559-1570.

Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck. 2001b. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. American Water Resources Assoc.* 37(5): 1169-1188.

Santhi, C., J. G. Arnold, R. Srinivasan, and J. R. Williams. 2003. A modeling approach to evaluate the impacts of water quality management plans implemented in the Big Cypress Creek watershed. In *Proc. Conf. Total Maximum Daily Load (TMDL) Environmental Regulations II*, 384-394. Ali Saleh, ed. ASABE Paper No. 701P1503. St. Joseph, Mich.: ASAE.

Schilling, K. E., and C. F. Wolter. 2009. Modeling nitrate-nitrogen load reduction strategies for the Des Moines River, Iowa, using SWAT. *Environ. Mgmt.* 44(4): 671-682

Searing, M. L., and A. Shirmohammadi. 1994. The design, construction, and analysis of a GIS database for use in reducing nonpoint-source pollution on an agricultural watershed. ASAE Paper No. 943551. St. Joseph, Mich.: ASAE.

Sexton, A. M., A. M. Sadeghi, X. Zhang, R. Srinivasan, and A. Shirmohammadi. 2010. Using NEXRAD and rain gauge precipitation data for hydrologic calibration of SWAT in a northeastern watershed. *Trans. ASABE* 53(5): 1501-1510.

Sexton, A. M., A. Shirmohammadi, A. M. Sadeghi, and H. J. Montas. 2011. Impact of parameter uncertainty on critical SWAT output simulations. *Trans. ASABE* 54(2): 461-471.

Shen, Z., Q. Hong, H. Yu, and R. Liu. 2008. Parameter uncertainty analysis of the nonpoint-source pollution in the Daning River watershed of the Three Gorges Reservoir region, China. *Sci. Total Environ.* 405(1-3): 195-205.

Shirmohammadi, A., and H. Montas. 2003. Project details of Warner Creek watershed: Section 319 national monitoring program project. College Park, Md.: University of Maryland, Department of Biological Resources Engineering. Available at: www.bre.umd.edu/projects/monocacy/main.htm. Accessed 2 July 2007.

Shirmohammadi, A., I. Chaubey, R. D. Harmel, D. D. Bosch, R. Muñoz-Carpena, C. Dharmasri, A. Sexton, M. Arabi, M. L. Wolfe, J. Frankenberger, C. Graff, and T. M. Sohrabi. 2006. Uncertainty in TMDL models. *Trans. ASABE* 49(4): 1033-1049.

Sohrabi, T. M., A. Shirmohammadi, T. W. Chu, H. Montas, and A. P. Nejadhashemi. 2003. Uncertainty analysis of hydrologic and water quality predictions for a small watershed using SWAT2000. *Environ. Forensics* 4(4): 229-238.

Tolson, B. A., and C. A. Shoemaker. 2007. Cannonsville reservoir watershed SWAT2000 model development, calibration, and validation. *J. Hydrol.* 337(1-2): 68-86.

USDA-SCS. 1990. Maryland Chesapeake Bay cooperative river basin study. Washington, D.C.: USDA Soil Conservation Service.

USEPA. 1997. Guidelines for preparation of the comprehensive state water quality assessments (305(b) reports) and electronic updates: Report contents. EPA-841-B-97-002A. Washington, D.C.: U.S. Environmental Protection Agency.

USEPA. 1999a. Draft guidance for water quality-based decisions: The TMDL process. 2nd ed. EPA 841-D-99-001. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.

USEPA. 1999b. Protocol for developing nutrient TMDLs. EPA 841-B-99-007. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.

USEPA. 2002a. The twenty needs report: How research can improve the TMDL program. EPA 841-B-02-002. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.

USEPA. 2002b. Consolidated assessment and listing methodology (CALM): Toward a compendium of best practices. 1st ed. Washington, D.C.: U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds.

USEPA. 2003a. Guidance for 2004 assessment, listing, and reporting requirements pursuant to sections 303(d) and 305(b) of the Clean Water Act. TMDL-0103. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.

USEPA. 2003b. National primary drinking water standards. EPA 816-F-03-016. Washington, D.C.: U.S. Environmental Protection Agency, National Service Center for Environmental Publications (NSCEP).

USEPA. 2007. State of the streams: 1995-1997 Maryland biological stream survey results. Annapolis, Md.: Maryland Department of Natural Resources. Available at: www.epa.gov/maia/html/mbss-ch10.html www.dnr.state.md.us/streams/pdfs/ea99-6.pdf. Accessed 7 June 2007.

Wagner, R. C., T. A. Dillaha, and G. Yagow. 2007. An assessment of the reference watershed approach for TMDLs with biological impairments. *Water Air Soil Pollut.* 181(1-4): 341-354.

Walker, W. W. 2003. Consideration of variability and uncertainty in phosphorus total maximum daily loads for lakes. *J. Water Resources Planning and Mgmt.* 129(4): 337-344.

White, K. L., and I. Chaubey. 2005. Sensitivity analysis, calibration, and validation for a multisite and multivariable SWAT model. *J. American Water Resources Assoc.* 41(5): 1077-1089.

Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, and H. Yang. 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe basin in China. *J. Hydrol.* 358(1-2): 1-23.

Zhang, H. X., and S. L. Yu. 2004. Applying the first-order error analysis in determining the margin of safety for total maximum daily load computations. *J. Environ. Eng.* 130(6): 664-673.