Enhanced HSPF Model Structure for Chesapeake Bay Watershed Simulation

Gary W. Shenk[1]; Jing Wu[2]; and Lewis C. Linker[3]

Abstract: For more than two decades, an HSPF-based watershed model has been used to simulate nutrient and sediment load delivery to the Chesapeake Bay. Over time, the watershed model has increased in complexity commensurate with the management challenges in Chesapeake Bay restoration. The increased complexity poses challenges to the standard application of HSPF for efficient operation of the model in a large-scale watershed, as well as difficulties in incorporating changes in best management practices (BMPs) and land uses over time. In response, the U.S. Environmental Protection Agency's Chesapeake Bay Program Office developed a software solution that enhances the existing HSPF model structure. The software system, consisting of preprocessors, an external transfer module, and postprocessors, was devised to conveniently generate and update parameter files essential to operations of a large and complex watershed-modeling system and to implement land-use and non-point-source-pollution management changes on any timescale greater than or equal to daily. The developed model system is demonstrated through comparison of the hydrologic calibrations of the current Phase 5 model and the previous Phase 4.3 model at 14 stations, as well as by several key scenario runs. The results show that the combined upgrades in segmentation, input data, and functionality improved model calibration; however, simply incorporating changes in land use did not significantly improve model calibration. The developed software provides a means to represent the key forcing functions in more detail and to address issues of flexibility that are difficult to manage in traditional HSPF applications. **DOI: 10.1061/(ASCE)EE.1943-7870.0000555.** *© 2012 American Society of Civil Engineers.*

CE database subject headings: Watersheds; Models; Chesapeake Bay; Nutrients; Sediment.

Author keywords: Watershed modeling; HSPF; Watershed management; Chesapeake Bay; TMDL.

Introduction

Effective control of water pollution at watershed scales is an important issue in the United States because water quality in many of the nation's water bodies, including the Chesapeake Bay, has become degraded due to excessive inputs of nutrients and sediment from non-point-source pollution. Increasingly, watershed-scale environmental models are being used to assist in managing non-point-source pollution (Tsihrintzis et al. 1996; Linker et al. 1996; Santhi et al. 2001; Borah and Bera 2004) and, more recently, in supporting development of total maximum daily load (TMDL) allocations (U.S. EPA 2001; Borah et al. 2006). In recognizing the importance of watershed modeling for TMDL development, the U.S. EPA has developed a geographic information system (GIS)–based software system called BASINS (Better Assessment Science Integrating Point and Nonpoint Sources; U.S. EPA 2001) to help carry out multipurpose watershed and water-quality-based analysis.

Watershed modeling has been an important component of the effort to understand non-point-source loading to the Chesapeake Bay and to develop management strategies for controlling it. For more than two decades, a watershed model, based on HSPF, has been used to simulate hydrology, nutrients, and sediment in the Chesapeake Bay watershed and to evaluate management options in reducing nutrient and sediment loads to achieve Chesapeake Bay restoration goals (Linker et al. 2000a). The results of model simulations were an integral part of the landmark *1987 Chesapeake Bay Agreement* (Baliles et al. 1987), several subsequent agreements (e.g., Gilmore et al. 2000), and most recently, the 2010 Chesapeake Bay TMDL (U.S. EPA 2010b).

Over time, the model has had many upgrades and refinements commensurate with the management challenges in Chesapeake Bay restoration (Linker et al. 2002). The current development version, the Phase 5 Community Watershed Model (the Phase 5 model), is a new generation of the watershed model developed in response to the needs of developing a large-scale assessment of the Chesapeake Bay TMDL with inputs and outputs on a smaller scale more relevant to management. The central organizing principle of the Phase 5 model is to simulate the land and river segments in separate input files and separate runs of HSPF, rather than within a single input file as is typically done. Land segments are defined in this paper as a set of land uses that are physically contiguous and share precipitation and meteorological data and generally coincide with county boundaries. Each land segment has a separate simulation of each of the 25 land-use types. All 25 land-use types are simulated whether or not they have existed within the land segment to allow for the possibility of incorporating them into future scenarios. River segments are the smallest units of river simulation in the Phase 5 model. Compared with the previous versions of model development, the Phase 5 model has a longer simulation period from 1984 to 2005; finer spatial segmentation of 1,063 river segments and 308 land segments; 25 land-use types, including several types of agricultural, developed, and forest land uses; and more detailed

---

[1] Integrated Analysis Coordinator, EPA Chesapeake Bay Program Office, 410 Severn Ave., Suite 112, Annapolis, MD 21403 (corresponding author). E-mail: GShenk@chesapeakebay.net

[2] Formerly, Associate Research Scientist, Univ. of Maryland Center for Environmental Science, 410 Severn Ave., Suite 112, Annapolis, MD 21403.

[3] Modeling Coordinator, EPA Chesapeake Bay Program Office, 410 Severn Ave., Suite 112, Annapolis, MD 21403.

Note. This manuscript was submitted on May 7, 2010; approved on February 17, 2012; published online on February 21, 2012. Discussion period open until February 1, 2013; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Environmental Engineering***, Vol. 138, No. 9, September 1, 2012. © ASCE, ISSN 0733-9372/2012/9-949-957/$25.00.

J. Environ. Eng. 2012.138:949-957.

input data for key loads such as atmospheric deposition (Grimm and Lynch 2005).

The scale and complexity of the Phase 5 model would pose a challenge to a standard application of HSPF. Given the scale of the Chesapeake Bay watershed, the number of input files that must be modified during calibration and scenario runs is large. The situation becomes further complicated with most river-segments receiving loads from multiple land segments, most land segments draining to multiple river segments, and more than 3 million individual applications of fertilizer or manure in the watershed that must be specified or changed for model runs. Creating all of these input files and land-water connections manually or in a graphical user interface environment such as BASINS is infeasible.

Another challenge is addressing changes in land use and best management practices (BMPs) during the 22-year simulation period. Over the two-decade calibration period, land uses and BMPs have gone through considerable changes. These changes affect the pollutant exports from the land and pollutant transport in streams. In the standard application of HSPF, these changes are difficult to simulate. The modules within HSPF that convert land export loads to river input loads have multiplication factors based on land-use acreages that are usually invariant through time. Similarly, factors that control changes in load attributable to BMPs are static as well. This lack of flexibility limits the model's ability to realistically mimic watershed situations over a long simulation period. The special actions module of HSPF could be used to change land-use acreages and BMP aggregate removal efficiencies throughout the model simulation period, but given the large number of input files that must be modified during calibration and scenario runs for the Phase 5 model, it is inefficient to utilize the special actions module to make these changes. Moreover, to use the special actions module for land-use and BMP changes, it would be necessary to have the land and river simulation within the same user control input (UCI). As discussed subsequently, the Phase 5 model is not constructed this way.

To date, HSPF has been widely reviewed and applied, especially after its inclusion into BASINS (Tsihrintzis et al. 1996; Love and Donigian 2002; Im et al. 2003; Filoso et al. 2004). Most of the applications are on relatively smaller basins than the six-state Chesapeake watershed (Borah and Bera 2004). As one of the largest applications of the HSPF model (Donigian et al. 1994), the Chesapeake Bay watershed model faces unique spatial- and temporal-scale issues. This paper presents a complementary system of software developed for HSPF to remove the limitations discussed previously. The software system aims to utilize the strengths of HSPF while providing the flexibility necessary for a large-scale watershed simulation over multidecadal simulation periods.

The EPA's BASINS software (U.S. EPA 2001) is an attempt to deal with some of the same issues. The BASINS software also automatically generates model input files through the use of parameters and specifications contained in databases or generated through GIS. When the Phase 5 model project began, BASINS did not support all of the HSPF functions and modules necessary for the Chesapeake Bay Program modeling and was considered insufficient as a possible platform. During the development of the Phase 5 model, BASINS was improved to include all HSPF functionality but was still considered to be less capable of handling a system as spatially complex as the 166,000-km^2 Chesapeake Bay Phase 5 model domain or providing the additional functionality of time-varying land-use and BMP-implementation acreages, without reworking the entire BASINS software. Recently, BASINS 4.0 has been released using exclusively open source software (U.S. EPA 2007). Had BASINS 4.0 been available earlier, it might have served as an appropriate platform on which to build the extensive programming structure developed for the Phase 5 model.

Hydrologic Simulation Program FORTRAN

Because it is a widely used watershed model, HSPF is in continual development and supported by several federal agencies (Bicknell et al. 1997). It is a continuous, conceptual, lumped-parameter model, which simulates hydrology, sediment, and chemical pollutants in the soil and streams. The model uses meteorological information, land surface characteristics, fertilizer and manure application data, and management practice information to simulate the processes that occur in a watershed. The HSPF model typically runs at an hourly time step and produces time series of flow, sediment load, and nutrient and pesticide loads at any segment in the watershed. An HSPF model is normally calibrated to observed flow and water-quality data measured at a river-segment outlet.

For simulation with HSPF, a basin is represented as a set of pervious or impervious unit areas and a set of river reaches and reservoirs that receive the output of the land simulation. Spatial variability of a watershed can be adequately represented by dividing the basin into many hydrologically homogeneous land segments and simulating runoff for each segment independently. Each simulated river reach or reservoir has a unique set of pervious and impervious land simulations that represent the land use within its watershed.

The land simulation module divides the soil into four vertical layers, with each layer having a separate but linked simulation of hydrologic, sediment, and nutrient processes. The land simulation includes detailed dynamics of nutrient balance and allows for detailed inputs of field operations and management actions through a special actions module. Outputs from a unique set of pervious and impervious land simulations that represent the land uses within a given subbasin can be targeted to a simulated river reach or reservoir associated with that subbasin. The hydrologic and water-quality processes that occur in the river channel network are simulated by a reach module. Water in a given reach is assumed to be completely mixed, and flows are routed to downstream reaches by kinematic-wave methods.

Phase 5 Model

The Chesapeake Bay watershed covers portions of six Middle Atlantic states (New York, Pennsylvania, West Virginia, Maryland, Delaware, and Virginia) and all of the District of Columbia (Fig. 1). The current land uses in the 166,000-km^2 Chesapeake Bay watershed are approximately 24% agriculture; 11% developed; and 65% forest, wooded, or other. At the request of Maryland and Virginia, the Phase 5 model domain includes all of the Chesapeake Bay watershed but also the entire states of Maryland, Virginia, and Delaware, plus portions of North Carolina and Tennessee, for an expanded model domain of 233,000 km^2. The portion of the model domain composed of counties that do not intersect the Chesapeake Bay watershed is available for hydrologic simulation only because data to fully support the nutrient and sediment simulation were not collected in those counties. This additional simulated area is used in the automatic hydrologic calibration to further constrain parameters.

The Phase 5 model divides the Chesapeake Bay watershed into 308 land segments, primarily on the basis of county boundaries, which is also the smallest unit of much of the available agricultural data (Martucci et al. 2005). The Phase 5 model has 25 land-uses

J. Environ. Eng. 2012.138:949-957.

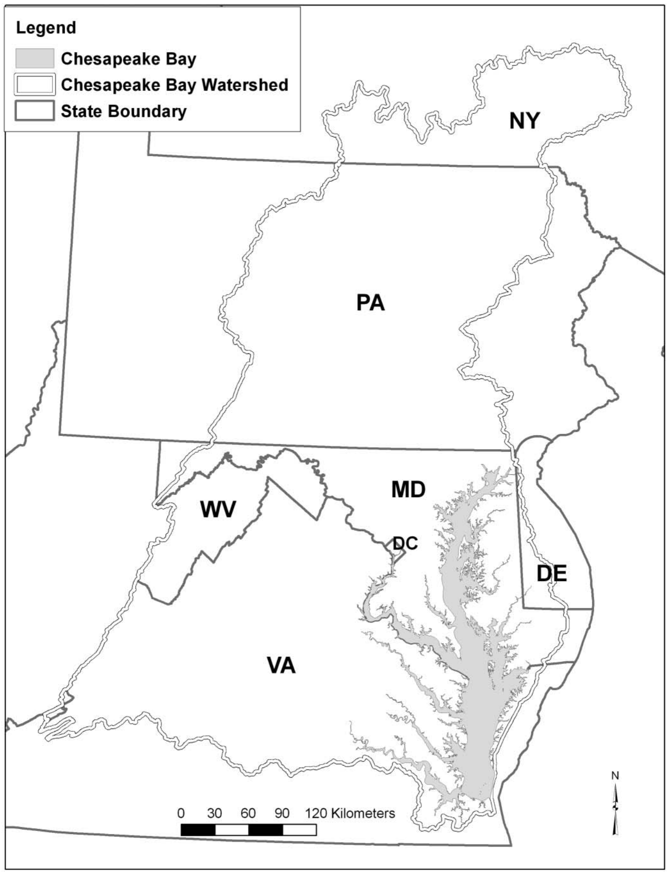

Fig. 1. Phase 5 model domain

types, including 11 types of cropland, 2 types of woodland, 3 types of pasture, 5 types of developed land, and provisions for other special land uses such as surface mines, animal-feeding operations, and land under construction. Each land-use type is simulated separately within each land segment. Nutrient processes in the major land uses of woodland, cropland, hay, pasture, and pervious urban are generally simulated using the AGCHEM module within HSPF, which simulates forest or crop nutrient cycling, including uptake by plants. Nutrient processes in the minor pervious land uses, which are harvested forest, land under construction, nurseries, surface mines, and degraded riparian pasture, are simulated through the PQUAL module, which represents nutrient export through monthly varying concentration coefficients. Impervious land uses are simulated through the IQUAL module, which uses accumulation and runoff coefficients to simulate nutrient and sediment export. Each AGCHEM land use is simulated on an hourly time step tracing the fate and transport of input nutrient loads from atmospheric deposition, fertilizers, animal manure, and point sources. Each land use is simulated as a single acre in each segment, and this single acre is then multiplied by the acreage of each land use draining to each river segment.

The Phase 5 model has 1,063 river segments at an average size of approximately 170 km^2. The river-segment delineation for the Chesapeake Bay watershed is based on the consistently applied criterion that all river reaches with an average flow of 2.8 m^3/s or larger are explicitly simulated, and, except as noted subsequently, river reaches with an average flow below 2.8 m^3/s are not explicitly simulated (Martucci et al. 2005). This work was largely based on the USGS Chesapeake Bay SPARROW model river reach segments (Preston and Brakebill 1999). The criterion of 2.8 m^3/s was relaxed to 1.4 m^3/s for the coastal plain and for river reaches that have observed streamflow data. Areas close to the tidal Chesapeake Bay that have streams with average flows less than 1.4 m^3/s or areas with streams that flow out of the study area were simulated as aggregations of land use only. The portion of the model domain outside of the Chesapeake Bay watershed was delineated using a 1:500,000-scale river reach network. The relatively fine spatial scale for the Phase 5 model allows for the inclusion of 287 flow calibration stations, representing an order of magnitude increase compared with the 20 stations used for the previous Phase 4.3 model. Increased segmentation improves characterization of spatial variation within the limitations of the "lumped-parameter" HSPF model.

Typically, in an application of HSPF, the land segmentation and the river segmentation cover the same area, with each river segment having a unique set of pervious and impervious land simulations. The decision to use separate segmentation for the land and river simulations within the Phase 5 model is based on several factors. The time to run the full watershed model with the current setup is a little less than 2 days on a single processor with the authors' current computers. Expanding the land segmentation to match the river segmentation would be prohibitive in terms of computer time. County-based land segmentation makes sense given that most of the important data sets on agricultural crop types and animal populations are not publicly available at levels smaller than a county, and that some of the Chesapeake Bay Program state partners are interested in the potential of using the output of the model to make environmental management decisions at the county level. Most counties are completely within a single broad geomorphic region, though some counties are broken into more than one segment to allow for orographically driven differential precipitation patterns. For a full discussion of the segmentation, see Martucci et al. (2005).

The model simulation period is 22 years, from 1984 to 2005, with the years 1985 to 2005 used as the calibration period, to take advantage of a wide range of monitored flow and loads of nutrients and sediment. The more detailed segmentation and more land-use types require 7,392 input files for independent land simulations, 930 input files for river simulations, and more than 45,000 land-use/river connections during the calibration and scenario runs. Also, more than 3 million individual nutrient applications of fertilizer or manures to different crops must be specified for the calibration or changed for a scenario run for the entire watershed during a 21-year simulation. Full documentation is available (U.S. EPA 2010a).

Enhanced Model Structure

In most HSPF applications, including simulations using WinHSPF (Duda et al. 2001) and BASINS, all land and river simulation modules are parameterized within a single UCI file. The water, nutrient, and sediment exports of each land use are multiplied by a single factor for land-use acreage and another factor for translation between land variable types and river variable types and units. Neither the land use nor the translation factors can be changed over simulation periods without significant effort through the special actions module.

An enhanced HSPF model structure was developed to provide for overall flexibility in model simulation. The core of the enhancement is to split the standard HSPF model structure into separate land and river simulations, allowing intermediate software to efficiently incorporate changes in land uses and management over time. Specifically, each land use in each land segment is parameterized in a separate UCI and is run in a separate invocation of HSPF. Similarly, each river segment has a separate UCI and invocation of HSPF. The enhanced software consists of preprocessors, an external transfer module (ETM), and postprocessors, which are used to generate input files for land and river simulations, link the land simulation to the river simulation, and compile and display model outputs, respectively (Shenk and Linker 2002). The auxiliary software, combined with a series of predeveloped databases and specification files, forms a complex and unique modeling system.

A convenient way to describe the Phase 5 model structure is through a step-by-step illustration of a model run. A typical Phase 5 model run consists of seven steps. Each step relies on the previous step(s) to receive required information: (1) identify upstream watershed; (2) generate land UCI files; (3) run HSPF land simulations; (4) convert land output to river input; (5) generate river UCI files; (6) run HSPF river simulations; and (7) compile model outputs for display, analysis, or input to a downstream model. A description of the software, functionality, and supporting files at each step is given subsequently. For full details, see U.S. EPA (2011).

Identify Upstream Watershed

Before a simulation is run for any particular watershed, upstream land segments and upstream river segments must be identified. The program BasinGen was developed to conveniently identify any subbasin of interest. BasinGen tracks the river network draining to any downstream segment and provides a list of land segments at least partially within the upstream watershed. The river network is generated on the basis of the river-segment naming convention detailed in Martucci et al. (2005). The connections between land segments and river segments are preprocessed through GIS tools and stored in an ASCII file for the entire model domain. The outputs of the BasinGen run are used in all subsequent steps to identify which land and river segments to run.

Generate Land UCI File

As noted previously, the Phase 5 model is structured so that each land-use type simulation within each land segment requires a unique UCI file. A program called Land UCI Generator (LUG) is designed to automatically generate UCI files for land simulations. To create a UCI file, the LUG does the following: (1) obtains operation instructions from a user-defined control file; (2) reads input data, parameter types, and parameter values from predeveloped databases; and (3) writes all information into a UCI file.

Before running the LUG, the user creates a land scenario control file that contains specifications to create UCI files for all land uses and land segments for a given scenario. The control file specifies for a particular scenario the simulation period, the HSPF modules that are active for each land-use type, the input data sets to be used, the set of parameter values to be used, and any nonstandard outputs to be generated. For example, a scenario that simulates only hydrologic processes on agriculture land will only need the activation of three HSPF modules, ATEMP, SNOW, and PWATER, and will not need to specify data sets for fertilizer and manure. In the normal scenario mode with a calibrated model, water-quality HSPF modules will be active, and scenario-specific data sets must be provided. The use of a single land scenario control file for all land uses and land segments allows users to control and manage model runs with great flexibility and efficiency.

The input data sets that can be specified for a given scenario are precipitation and meteorological data (air temperature, dew point, wind speed and direction, solar radiation, cloud cover, and potential evapotranspiration), vegetative cover, fertilizer and manure applications, legume fixation, potential nutrient uptake by crops, and atmospheric deposition of nutrients. These data sets are preprocessed either by spreadsheet or database analysis or specialized programs and stored in a group of ASCII files that are formatted in accordance with the read/write functionality of the LUG.

For a typical Phase 5 model run at the Chesapeake Bay Program Office, management-related inputs are supplied by a database application known as Scenario Builder (U.S. EPA 2010d).

As nutrient application rates and other crop-related information change over time, several data sets for each data type can be specified with a date for which each one is applicable. The LUG interpolates and extrapolates linearly for all other dates. This functionality is typically used only in calibration mode.

Run HSPF Land Simulation

Once a UCI file for each land use within each land segment is generated, HSPF is run on each land UCI. For each land use and each land segment, the simulated flow and pollutant time series are stored in individual watershed data management (WDM) files, the most efficient method of input/output for HSPF.

Convert Land Output to River Input

An ETM was developed to connect the land simulation to the river simulation. The ETM consists of a group of routines that direct the appropriate water volume and nutrients and sediment loads from each land-use type within each land segment to each river segment, and it is the core part of the Phase 5 model enhancement. The ETM performs the following broad functions: (1) multiply land output by appropriate coefficients to account for units and type conversions, (2) multiply unit-area land outputs by acreages reflecting the area of a land use within a land segment that contributes to a river segment, and (3) account for changes in loads attributable to management practices.

A typical unit conversion would be to divide the land-water output unit of acre-inches by 12 to convert to the river water input unit of acre-feet. A type conversion would be to divide the single sediment type class simulated from a land segment into the sand, silt, and clay components that are simulated in a river segment. These relations may be specified globally or by segment.

Land-use data are available for a 5-year interval, for the years 1982, 1987, 1992, 1997 and 2002, derived from the land cover data from multiple sources, the Agricultural Census data from USDA, and ancillary data. For a full description, see U.S. EPA (2010a, section 4). The resultant land-use data are stored in tabular databases, representing one point in time. The ETM is programmed to accept land-use data at several points in time and to interpolate and extrapolate these data linearly through time as necessary. Land use is allowed to vary for calibration but is held constant for scenarios.

The BMP simulation within the ETM accounts for (1) the effectiveness of each BMP type on each land use, (2) the change in the number of acres that are affected or managed by the BMP through time, (3) the effect of multiple BMPs at the same physical location, (4) the effect of hydrology on BMP effectiveness, and (5) an optional randomizing effect. These features are briefly described subsequently.

The nominal effectiveness, or removal efficiency, of each BMP type on each land use for each constituent is stored in a database. Nominal effectiveness values are largely based on Simpson and Weammert (2009), but are frequently updated by expert panels associated with the Chesapeake Bay Program. The BMP effectiveness database may be modified for each scenario to accommodate new BMPs or assess sensitivity to effectiveness assumptions.

The aggregate effect of a BMP on a land-use type is dynamically simulated by multiplying the nominal effectiveness by a time series produced by dividing the managed acres under a given BMP by the total acres of that land use. This area-weighted effectiveness is reduced for the possibility of spatially overlapping BMPs by multiplying the complements, or 1 minus the values, of area-weighted effectiveness values for BMPs that may possibly overlap. Full descriptions are available in U.S. EPA (2010a, section 6).

The effect of hydrology on BMP effectiveness is taken into account by introducing an hourly factor that reduces the effectiveness as a function of the return frequency of runoff for each hour. The relationship between return frequency and effectiveness reduction can be specified separately for each BMP within a predefined set of functions. A random effect is also available. The user may specify a distribution type within a predefined set and associated parameter values. The random effect is not desirable for management scenarios and is typically not used.

Generate River UCI File

A program called River UCI Generator (RUG) was developed to provide the functionality of generating river UCI files. In a similar fashion to the LUG, the RUG (1) obtains operation instructions from a user-defined control file; (2) reads input data, parameter types, and parameter values from predeveloped databases; and (3) writes all information into a UCI file.

As with the LUG, the user creates a river scenario control file for the RUG, which identifies the simulation period, the active HSPF modules, the land scenario to run for each land-use type, the input data sets to be used, the set of parameter values to be used, and any nonstandard outputs to be generated.

The input data sets that can be specified for a given scenario are precipitation and meteorological data, land use, BMP managed acres, point sources, septic loads, and water diversions. As with the LUG, Scenario Builder (U.S. EPA 2010d) normally supplies these inputs.

Run HSPF River Simulation

HSPF is run on each river UCI, and the outputs of river simulation are written back to each river WDM. At this step, a program is also run to combine loads from point sources, septic systems, and atmospheric deposition into each river WDM, as well as subtract any water diverted from a river. The data sets for point sources, septic systems, atmospheric deposition, and water diversion are preprocessed from data for the entire Chesapeake Bay watershed and stored in individual WDMs on the basis of their locations in the watershed. Putting these data sets into separate WDMs and invoking them at this stage allow users to easily make changes in these data sets without having to rerun any previous steps, another feature designed to provide flexibility and efficiency for model runs.

Compile Model Outputs

The model simulation produces a large body of information that is stored in land and river WDMs and in binary files with additional information on loads known as "transfer coefficient binaries." To assist in model analysis, a large group of postprocessors are developed to transfer model results from WDMs to ASCII files for display or further analysis. For any basin in the watershed, the postprocessors can summarize outputs (flow and loads) at various temporal scales (daily, monthly, annual, and average annual), compute relevant statistics, compare simulated to observed values, or compile information for input to a downstream model. The functions of the postprocessors are all put in a single script, and each of them is controlled by an independent program, thereby allowing users to easily specify any output of interest.

Overall Functionality of the Software System

To summarize the previous steps, the overall model structure and a typical application process are illustrated in Fig. 2. Separating land and river simulation into different UCIs provides great flexibility in model simulation. With this structure, each land-use type simulation within each land segment is completely independent of any other land or river simulation, and each river simulation is dependent only on the local land-use type simulations and the upstream river simulations. This provides an efficient way to deal with the complicated land-river/river-river logistics that challenge a large-scale watershed simulation.

Specifically, the developed software system has several advantages over a traditional HSPF application: (1) it easily allows for large-scale parameter-value adjustments during calibration; (2) parallel computing operations become convenient, and thus simulation can be arranged more efficiently; (3) addition of new land-use types is relatively easy; (4) integration into outside databases for scenarios is convenient; and (5) it can be run for any watershed within the larger model domain.

The model system is run on personal and server-class computers with the Linux operating system. All supporting programs, as well as HSPF, are open source and written primarily in FORTRAN 77.

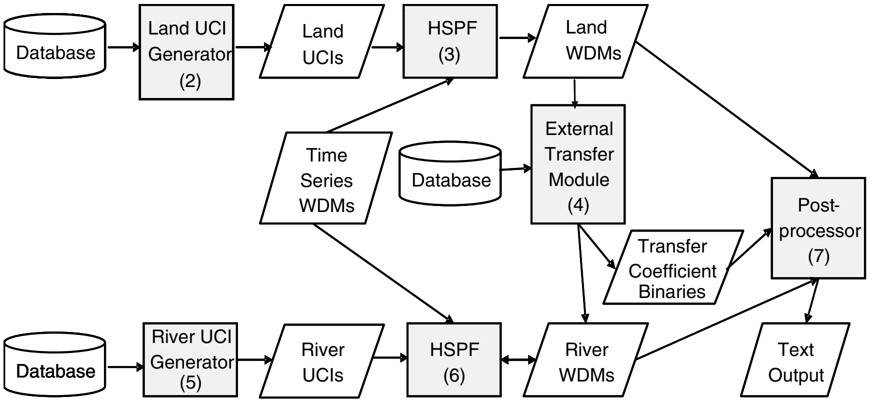

Fig. 2. Information flow of the developed model structure; numbers in parentheses refer to the steps described in the text

Most operations are controlled by *c*-shell Linux scripts. The Chesapeake Bay Program Office runs the Phase 5 model on a 32-processor server, reducing run time from more than 2 days to a few hours.

Application of Developed Model System

A full description of the Phase 5 model calibration methods and results and scenario runs is beyond the scope of this paper. The following paragraphs describe a comparison between the Phase 5 model and the previous management model of the Chesapeake Bay watershed to show the effect of advances in scale and structure, show an example of the use of the Phase 5 modeling system to gain insight into the importance of using time-varying land-use data, and describe selected scenario runs of critical management strategies to demonstrate the effectiveness and efficiency of the model structure in running scenarios.

Hydrologic Calibration Improvement

The Phase 5 model was calibrated at 287 flow stations across the entire Chesapeake Bay watershed (Fig. 3). The calibration period is from 1985 to 2005.

The hydrologic calibration was performed using an automated method in which model parameter values were updated on the basis of a heuristic system of rules that relate model parameter-value adjustments to statistical abstractions of the simulated and observed hydrographs. The Nash-Sutcliffe model efficiency (Beven 2001) was used as a summary statistic and was considered particularly useful as such because it was not used directly in the calibration.

The combined effects of the improvements in spatial segmentation, input data, calibration method, and model structure are demonstrated through the comparison of the Phase 5 model hydrologic calibration with the previous Phase 4.3 model calibration at 14 stations (Fig. 3). Detailed information regarding model input data and calibration rules for the Phase 4.3 model can be found in Linker et al. (2000b). The model efficiency for hydrologic calibration shows that the Phase 5 model is better calibrated to observed data than the Phase 4.3 model for all stations except one (Table 1). The average increase in efficiency from Phase 4.3 to Phase 5 is 0.28.

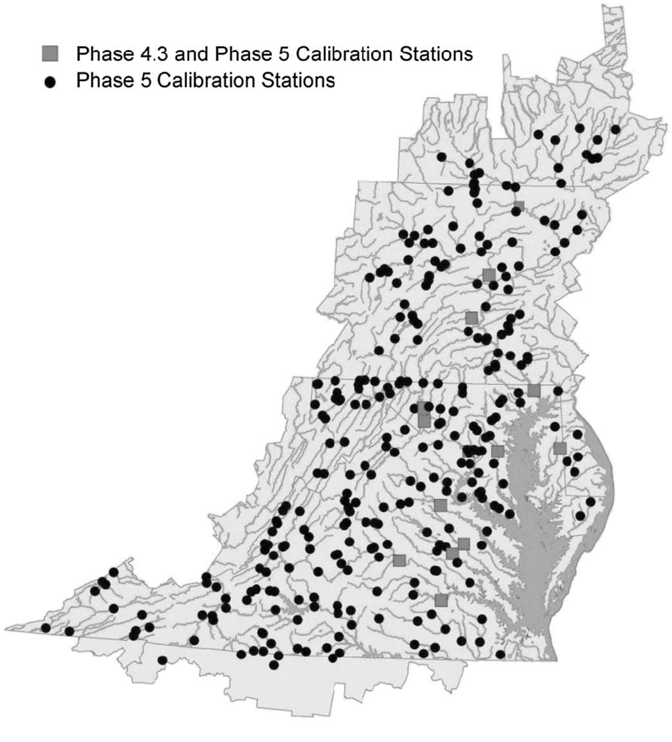

Fig. 3. Calibration stations in Chesapeake Bay watershed

Test of the Importance of Time-Varying Land-Use Data

During the calibration period of 1985 to 2005, agriculture decreased by an estimated 11% and developed land increased by an estimated 18% in the Chesapeake Bay watershed (U.S. EPA 2010a). As a test of the role of land-use change in the calibration improvements, a separate hydrologic calibration was performed with constant 1995 land use instead of the time-varying land use. The results for model efficiency were nearly identical at individual stations, and the distributions of the model efficiencies for the 287 stations were statistically indistinguishable between the two calibrations. In this case, using a static land use from the midpoint of the calibration period produced no decrease in calibration accuracy versus a model with time-varying land use. The effort to create the time-varying land-use data set was still useful in that scenarios require different land-use years as baselines.

Scenario Runs

The calibrated watershed model was used to develop and test different management scenarios related to the effort to reduce nutrients and sediments to achieve water-quality standards of dissolved oxygen, chlorophyll, and water clarity (U.S. EPA 2010b). A series of scenarios was designed to estimate the nutrient and sediment load reductions associated with increased implementation levels of BMPs, wastewater-treatment upgrades, and/or other point or non-point control technologies (U.S. EPA 2010c). Of particular policy importance were (1) the 2009 scenario, representing current BMP-implementation levels throughout the watershed; (2) the tributary strategy scenario, representing river-specific cleanup strategies that detail the "on-the-ground" actions needed to meet load reduction goals; and (3) the E3 scenario (everything, everywhere by everybody), representing the *maximum* theoretical implementation of the *best combination* of BMPs or other control technologies available to a land use, point source, or other load source (U.S. EPA 2003). The 2009 scenario was based on 2009 data, whereas the tributary strategy and E3 scenarios were based on 2010 projections of land uses, human populations, agricultural animal populations, point-source flows, septic systems, and atmospheric deposition.

The input data sets were the same for all scenarios except for management practice data as described in U.S. EPA (2010c). The model was then run for each scenario, following the seven steps described in the previous section. The functionality of the Phase 5 model structure made the scenario runs relatively fast and efficient.

The scenario runs estimated that total nitrogen loads were 114 million kg for the 2009 scenario, 86.7 million kg for the tributary strategy scenario, and 64.0 million kg for the E3 scenario. Total phosphorus loads for the three scenarios are 7.5 million, 6.5 million, and 3.9 million kg, respectively (U.S. EPA 2010c). These scenario runs and others provided a scientific basis for deriving state-basin target load allocations, and in helping guide development of the state TMDL watershed implementation plans. Ultimately, the Chesapeake Bay TMDL allocations were 84 million kg of nitrogen, 5.7 million kg of phosphorus, and 29 million kg of sediment.

Table 1. Model Efficiencies of Hydrologic Calibration of the Phase 4.3 and Phase 5 Models

Location	Phase 5 river	Phase 4.3	Phase 5
Choptank River near Greensboro, Maryland	EM2_3980_0001	0.45	0.68
Appomattox River at Matoaca, Virginia	JA5_7480_0001	0.56	0.71
James River at Catersville, Virginia	JL7_7100_7030	0.44	0.83
Potomac River near Washington, D.C.	PM7_4820_0001	0.46	0.8
Shenandoah River at Millville, West Virginia	PS5_4380_4370	0.5	0.82
Potomac River at Shepherdstown, West Virginia	PU6_3752_4080	0.56	0.81
Rappahannock River near Fredericksburg, West Virginia	RU5_6030_0001	−0.02	0.63
Juniata River at Newport, Pennsylvania	SJ6_2130_0003	0.58	0.82
Susquehanna River at Conowingo, Maryland	SL9_2720_0001	0.77	0.84
Susquehanna River at Towanda, Pennsylvania	SU7_0850_0730	0.65	0.84
West branch Susquehanna River at Lewisburg, Pennsylvania	SW7_1640_0003	0.6	0.8
Patuxent River near Bowie, Maryland	XU3_4650_0001	0.04	0.7
Mattaponi River near Beulahville, Virginia	YM4_6620_0003	0.52	0.37
Pamunkey River near Hanover, Virginia	YP4_6720_6750	0.45	0.82

The efficiency gains of the Phase 5 model structure were critical for the production of the final Chesapeake Bay TMDL on schedule. During the second half of 2010, the Chesapeake Bay Program partnership ran more than 150 management scenarios, including more than 70 watershed implementation plan scenarios.

Discussion

The intended purpose of the Phase 5 model is to estimate the effects of changes in nutrient inputs, management actions, and land use on the loads of nutrients and sediment delivered to the tidal Chesapeake Bay. During the 21-year calibration period of 1985–2005, the three key forcing functions of land use, management actions, and nutrient inputs have changed considerably. The Phase 5 model structure allows these time-varying forcing functions to be more accurately represented in the model simulation, resulting in more realistic inputs as a basis for calibration. For example, the handling of the management practice effectiveness as a function of rainfall greatly enhances the model's ability to correctly credit management actions taking place in the watershed. The overall system is an efficient and logical approach to dealing with the complex land-river and river-river logistics that challenge any large regional watershed-modeling program working at relatively fine spatial scales.

A comparison of calibration results with the prior model for the same watershed showed that the overall system changes, which included improvements in segmentation, input data, model structure, and calibration method, produced significantly improved results. However, the more realistic representation of changes in land use and management practices through time did not result in a better or worse hydrologic calibration.

With the proliferation of inexpensive Linux systems that can be clustered together, computing power that is normally reserved for large simulation projects is now affordable and generally accessible to more users. Because all land uses in a watershed model of this type are independent of one another, there exists a significant opportunity to take advantage of parallel computing. The Phase 5 model, with one land use or river reach per file, allows for a high degree of parallelization.

Calibration and scenario run times are reduced significantly. During the calibration of a watershed model with several land uses, land uses are typically calibrated first, followed by the riverine portions of the model. Separating the land simulation from the river simulation allows the user to store the calibrated land simulation in WDM files, and then only the river module would need to be run for the river calibration. Scenario run time can also be reduced if land-use simulations have the same precipitation, meteorological data, and nutrient applications as a previous scenario. For example, if a particular scenario only involves a change in land-use acreage, the land simulation does not need to be rerun, saving approximately half of the normal scenario run time.

Another benefit of the modularity described in this paper is the ability to integrate models other than HSPF into the system. If it were found that another stream model was more appropriate for a particular application, routines could be written to provide input in the necessary format. Provision was made for specialized land-use models, such as those for wetland, riparian forest buffer, or more detailed forest simulation, to be included. The Phase 5 model system could also provide an interface to other modeling or optimization frameworks.

The Phase 5 model structure is particularly suitable for serving as a community model. A community model consists of open source, public domain of model code, preprocessors, postprocessors, and input data that are freely distributed over the web. With the Phase 5 model system, the Chesapeake Bay Phase 5 model may be used in a direct as-is application or as a point of departure for a more detailed, small-scale model. Local watershed managers could make use of additional modeling tools and more site-specific local information to resegment, recalibrate, and implement the model at appropriate local scales. The use of the community model approach can ensure regional consistency of water-quality analysis and TMDL development, as well as consistency of local TMDLs with the large-scale regional TMDL of the Chesapeake Bay. This benefit should provide opportunities for more effective, cost-efficient, and equitable water-quality management.

Conclusion

The modeling system developed for the Chesapeake Bay Phase 5 Community Watershed Model is a versatile method that utilizes the strengths of HSPF and incorporates more time-dependent information than would be possible in a standard application. The Land UCI Generator and River UCI Generator allow the generation and modification of large numbers of input files in a convenient format, which is essential to calibration and scenario operations in a large and complex modeling system. The external transfer module allows for the opportunity to simulate a watershed over an extended period by providing a method to change land-use acreages and BMP-affected acreages and removal efficiencies over time.

J. Environ. Eng. 2012.138:949-957.

The software system allows the Phase 5 model to increase spatial segmentation by an order of magnitude from previous efforts, while maintaining the ability to administer the model efficiently and simulate the effects of land-use and management changes through time. The enhanced model structure provides a means to achieve a more accurate and efficient HSPF simulation for any large-scale watershed model.

The effectiveness and efficiency of the developed model system are demonstrated through comparison of the hydrologic calibrations of the Phase 5 model with the previous Phase 4.3 model and through the development of several scenario runs. The Phase 5 model is better calibrated to the observed data than the previous Phase 4.3 model for most cases, indicating that the combined upgrades in segmentation, input data, model functionality, and calibration method improves model calibration accuracy. However, the more realistic representation of changes in land use and management practices through time did not result in a better or worse hydrologic calibration. The Phase 5 model is an efficient tool for scenario operations as shown through rapid delivery of scenarios related to the Chesapeake Bay TMDL. The Phase 5 model is an open source, public domain model (Note: Several versions of the Phase 5 model have been developed and have been available for download. This work is based on the Phase 5.3.0 model used in the 2010 Chesapeake TMDL. As of this writing, the current version is Phase 5.3.2.). The Phase 5 model simulation system is made available, along with the entire model code, data library, and documentation, at http://ches.communitymodeling.org/models/CBPhase5/index.php.

References

Baliles, G. L., Schaefer, W. D., Casey, R. P., Thomas, L. M., Barry, M. S., and Cole, K. J. (1987). "1987 Chesapeake Bay agreement." Chesapeake Bay Program Executive Council, ⟨http://www.chesapeakebay.net/content/publications/cbp_12510.pdf⟩ (Jul. 6, 2012).

Beven, K. J. (2001). *Rainfall-runoff modelling: The primer*, John Wiley & Sons, Chichester, England, 360.

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Jr., Donigian, A. S., Jr., and Johanson, R. C. (1997). "Hydrological simulation program—FORTRAN, user's manual for version 11." *EPA/600/R-97/080*, U.S. Environmental Protection Agency, National Exposure Research Laboratory, Athens, GA.

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

Borah, D., et al. (2006). "Sediment and nutrient modeling for TMDL development and implementation." *Trans. ASAE*, 49(4), 967–986.

Donigian, A. S., Jr., Bicknell, B. R., Patwardhan, A. S., Linker, L. C., Chang, C. H., and Reynolds, R. (1994). *Chesapeake Bay Program watershed model application to calculate bay nutrient loadings*, U.S. EPA Chesapeake Bay Program Office, Annapolis, MD.

Duda, P., Kittle, J., Jr., Gray, M., Hummel, P., and Dusenbury, R. (2001). WinHSPF version 2.0, an interactive Windows interface to HSPF (WinHSPF), user's manual, *Office of Science and Technology, Office of Water*, U.S. EPA, Washington, D.C.

Filoso, S., Vallino, J., Hopkinson, C., Rastetter, E., and Claessens, L. (2004). "Modeling nitrogen transport in the Ipswich River basin, Massachusetts, using a Hydrologic Simulation Program in FORTRAN (HSPF)." *J. Am. Water Resour. Assoc.*, 40(5), 1365–1384.

Gilmore, J. S., Glendenning, P. N., Ridge, T. J., Williams, A. A., Browner, C. M., and Billing, W. T. (2000). "Chesapeake 2000." Chesapeake Bay Program Executive Council, ⟨http://www.chesapeakebay.net/documents/cbp_12081.pdf⟩ (Jul. 6, 2012).

Grimm, J. W., and Lynch, J. A. (2005). "Improved daily precipitation nitrate and ammonium concentration models for the Chesapeake Bay watershed." *Environ. Pollut.*, 135(3), 445–455.

Im, S., Brannan, K. M., and Mostaghimi, S. (2003). "Simulating hydrologic and water quality impacts of an urbanized watershed." *J. Am. Water Resour. Assoc.*, 39(6), 1465–1479.

Linker, L., Shenk, G., Dennis, R., and Sweeney, J. (2000a). "Cross-media models of the Chesapeake Bay watershed and airshed." *Water Qual. Ecosyst. Model.*, 1(1–4), 91–122.

Linker, L., Shenk, G., and Hopkins, K. (2000b). "Watershed model Phase 4.3 calibration rules." ⟨http://www.chesapeakebay.net/content/publications/cbp_13352.pdf⟩ (Sept. 26, 2011).

Linker, L., Shenk, G., Wang, P., Hopkins, K., and Pokharel, S. (2002). "A short history of Chesapeake Bay modeling and the next generation of watershed and estuarine models." *Proc.,Water Environment Federation, Watershed 2002*, Water Environment Federation, Alexandria, VA, Session 6, 569–582.

Linker, L., Stigall, C. G., Chang, C. H., and Donigian, A. S., Jr. (1996). "Aquatic accounting: Chesapeake Bay watershed model quantifies nutrient loads." *Water Environ. Technol.*, 8(1), 48–52.

Love, J., and Donigian, A. S., Jr. (2002). "The Connecticut watershed model—model development, calibration and validation." *Watershed 2002 (CD-ROM)*, Water Environment Federation, Alexandria, VA.

Martucci, S., Krstolic, J., Raffensperger, J., and Hopkins, K. (2005). "Development of land segmentation, stream-reach network, and watersheds in support of Hydrological Simulation Program—Fortran (HSPF) modeling, Chesapeake Bay watershed, and adjacent parts of Maryland, Delaware, and Virginia." *USGS SIR 2005-5073*, U.S. Geological Survey, Reston, VA.

Preston, S. D., and Brakebill, J. W. (1999). "Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed." *Water-Resources. Investigations Rep. 99-4054*, U.S. Geological Survey, Baltimore, MD, 12.

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

Shenk, G., and Linker, L. (2002). "Simulating the Chesapeake Bay watershed with time varying land use and management actions." *Proc., Water Environment Federation, Watershed 2002*, Water Environment Federation, Alexandria, VA, Session 2, 225–237.

Simpson, T. W., and Weammert, S. E. (2009). "Developing best management practice definitions and effectiveness estimates for nitrogen, phosphorus and sediment in the Chesapeake Bay watershed." ⟨http://archive.chesapeakebay.net/pubs/BMP_ASSESSMENT_REPORT.pdf⟩ (Sept. 26, 2011).

Tsihrintzis, V. A., Fuentes, H. R., and Gadipudi, R. (1996). "Modeling prevention alternatives for nonpoint source pollution at a well field in Florida." *J. Am. Water Resour. Assoc.*, 32(2), 317–331.

U.S. EPA. (2001). "Better Assessment Science Integrating Point and Nonpoint Sources—BASINS version 3.0 user's manual." *EPA-823-8-01-001*, EPA Office of Water, ⟨http://www.epa.gov/waterscience/basins/bsnsdocs.html⟩ (Sept. 26, 2011).

U.S. EPA. (2003). "Setting and allocating the Chesapeake Bay basin nutrient and sediment loads." *EPA 903-R-03-007*, EPA Region III, Chesapeake Bay Program Office.

U.S. EPA. (2007) "BASINS 4.0." EPA Office of Water, ⟨http://water.epa.gov/scitech/datait/models/basins/BASINS4_index.cfm⟩ (Oct. 1, 2011).

U.S. EPA. (2010a). "Chesapeake Bay Phase 5.3 Community Watershed Model." *EPA 903S10002-CBP/TRS-303-10*, ⟨http://ches.communitymodeling.org/models/CBPhase5/documentation.php⟩ (Oct. 1, 2011).

U.S. EPA. (2010b). "Chesapeake Bay total maximum daily load for nitrogen, phosphorus and sediment." ⟨http://www.epa.gov/reg3wapd/tmdl/ChesapeakeBay/tmdlexec.html⟩ (Oct. 1, 2011).

U.S. EPA. (2010c). "Chesapeake Bay total maximum daily load for nitrogen, phosphorus and sediment: Appendix J Key Chesapeake Bay TMDL reference and management modeling scenarios: Definitions and descriptions." ⟨http://www.epa.gov/reg3wapd/pdf/pdf_chesbay/FinalBayTMDL/AppendixJScenarios_final.pdf⟩ (Oct. 1, 2011).

U.S. EPA. (2010d) "Estimates of county-level nitrogen and phosphorus data for use in modeling pollutant reduction. Documentation for Scenario Builder version 2.2." EPA Chesapeake Bay Program Office, ⟨http://archive.chesapeakebay.net/pubs/SB_V22_Final_12_31_2010.pdf⟩ (Oct. 1, 2011).

U.S. EPA. (2011). "User's manual for the Chesapeake Bay Program's Phase 5 watershed model." EPA Chesapeake Bay Program Office, ⟨ftp://ftp.chesapeakebay.net/Modeling/phase5/community/Working_Copy_of_Manual/Model_Operations_Manual_p5_3_0.doc⟩ (Oct. 18, 2011).

J. Environ. Eng. 2012.138:949-957.