Using multiple watershed models to assess the water quality impacts of alternate land development scenarios for a small community

Amirreza Sharifi[1], Haw Yen[2†], Kathleen M.B. Boomer[3,4], Latif Kalin[5], Xuyong Li[3,6], Donald E. Weller[3]

[1] Department of Environmental Science and Technology, University of Maryland, College Park, Maryland 20742, USA

[2] Blackland Research and Extension Center, Texas A&M Agrilife Research, 720 East Blackland Road, Temple, Texas 76502, USA

[3] Smithsonian Environmental Research Center, Edgewater, Maryland 21037, USA

[4] The Nature Conservancy, Bethesda, Maryland 20814, USA

[5] School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA

[6] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China

[†] Corresponding author; email: haw.yen@gmail.com , hyen@brc.tamus.edu;
Tel: (254) 774-6004

Abstract

Chesapeake Bay, the largest estuary in North America, is impaired by excess nutrient discharges, especially from urban and agricultural land. Watershed simulation models have provided key insights to understanding land-to-water connections, but rarely are these models applied to guide local land management to explore and communicate uncertainty in the model predictions. In this study, three watershed simulation models; the Soil and Water Assessment Tool (SWAT), the Generalized Watershed Loading Function (GWLF) model, and the Chesapeake Bay Program's Chesapeake Watershed Model (CBP-CWM) were implemented to predict water, total nitrogen, and total phosphorus discharges from small tributaries in the town of Queenstown, Maryland, USA. Based on our evaluation metrics, none of the models consistently provided better results. In general, there was a good agreement on annual average water flow between the SWAT and CBP-CWM models, and the GWLF and CBP-CWM models predicted similar TN and TP loads. Each model has strengths and weaknesses in flow and nutrient predictions, and predictions differed among models even when models were initialized with the same data. Using multiple models may enhance the quality of model predictions and the decision making process. However, it could also be the case that the complexity of implemented watershed models and resolution of our understanding currently are not yet suited to provide scientifically credible solutions.

Keywords: Watershed Modeling; Multiple Model Comparisons; SWAT; GWLF; CBP-CWM; Chesapeake Bay

1. Introduction

Coastal zones provide valuable ecosystem services to human society worldwide (Agardy and Alder, 2005; Barbier et al., 2011), but coastal zones have also been foci of urban development. In some US coastal areas, the rate of development has considerably exceeded the population growth rate (Nagy et al., 2012). Population growth is accompanied by land conversion, mostly into urban land uses, which can threaten the integrity of coastal waters through multiple negative effects on water quality (Grimm et al., 2008; Tu, 2009). Urbanization increases impervious area, resulting in quicker and larger pulses in storm flow, geomorphic changes in stream channels, and higher sediment yields (Arnold et al., 1982; Wahl et al., 1997). Urban lands are also potential sources for heavy metals, nutrients, and bacteria (Rose, 2002; Schoonover et al., 2005). Excessive loads of nitrogen (N) and phosphorus (P) in urban streams can trigger undesirable effects in the receiving water bodies, such as algal blooms, eutrophication, and hypoxia. In addition to urbanization, agricultural activities are also major contributors to coastal eutrophication (Boesch et al., 2001).

Chesapeake Bay, the largest estuary in North America, is ecologically degraded, largely because of excessive nutrients received from urban and agricultural discharges. In 1970, Chesapeake Bay was one of the first estuaries found to contain marine dead zones (Kemp et al., 2005). The Bay and its tidal tributaries were later listed as impaired water bodies under section 303(d) of the Clean Water Act. Since 1980, management efforts to reduce nutrient loads to the Bay have intensified, but the loads from urban land have actually increased by 15% since 1985 (Chesapeake Bay Program, 2010). Increased loads from population growth and new suburban sprawl have outweighed load reductions achieved from stormwater management practices. Current efforts to reduce urban loads emphasize site-scale practices (i. e., stormwater management) and watershed-scale planning, such as directing low impact development to designated areas adjacent to a municipality.

Since 1983, the Chesapeake Bay Program (CBP); a regional partnership including local, state, and federal agencies,; has worked to protect and restore the Bay and its 167,000 km^2 watershed (Chesapeake Bay Program, 2010). To develop policy recommendations, the CBP uses simulation models of the Chesapeake Bay watershed (CBP-CWM) and estuary to set the regulatory limits for total maximum daily loads (TMDLs) to Chesapeake Bay and to evaluate the likely effects of possible management actions on nutrient loads (Linker et al., 2013). However, land management plans are implemented at much smaller spatial units than those considered by the CBP-CWM model. Furthermore, when assessing the impacts of alternative land management plans, the intrinsic uncertainty of watershed processes modeling and the potential impacts of climate change on surface water quality and quantity are often overlooked. Land management plans for improving water quality may fail if the plans are based on models that do not consider the spatial patterns of land use, model uncertainty, or climatic variability (Weller et al. 2011, Weller and Baker 2014).

Watershed models are essential tools for summarizing knowledge of watershed processes and forecasting the effects of different land use or climate scenarios on water quantity and quality. However, imperfect model representations of key hydrologic and biogeochemical processes reduce confidence in model predictions (Sharifi et al., 2016; Yen et al., 2014b). Combining results from a group of models (ensemble modeling) instead of relying on a single model can improve predictions and enhance confidence when applying the models to identify optimal development scenarios (Beven and Freer, 2001; McIntyre et al., 2005). Assessing model structural uncertainty is a common objective among many studies that have employed multiple watershed models (Breuer et al., 2009). Most of these studies focused only on parameter uncertainty within a single model, without much consideration to structural uncertainty (i. e., the choice of underlying model algorithms) or input uncertainty (i. e., the choice of and errors in land use, land cover, and other input data) (Vrugt et al 2005). Furthermore, most studies focus primarily on flow prediction (Reed et al., 2004; Goswami et al., 2005; Breuer et al., 2009); and fewer studies considered model uncertainty in predicting sediment (Kalin and Hantush, 2006; Shen et al., 2009), phosphorus (Nasr et al., 2007) nitrogen (Amiri and Nakane, 2009; Grizzettia et al., 2005), or multiple materials (Boomer et al. 2013).

A multi-model ensemble (MME) goes beyond model comparison by integrating the predictions of individual models into an ensemble average. MME often has better average performance than single models and increases the credibility of model predictions by accounting for uncertainty in model structure (Georgakakos et al., 2004; Boomer et al 2013). Ensemble model averaging provides alternatives in addition to a single model, especially when there is not enough information to identify the best model or when the data do not favor a particular model (Kadane and Lazar 2004). Several studies have applied the MME approach to flow prediction or flood forecasting (Renner et al., 2009; Zhao et al., 2011) and one study demonstrated that combining nitrogen predictions of five models gave better predictions than the individual models (Exbrayat et al., 2010). In addition, the LUCHEM study applied an ensemble of 10 watershed models to assess the effects of land use and land cover (LULC) change on hydrology and water quality (Breuer et al., 2009; Huisman et al., 2009; Viney et al., 2009).

It was mentioned in literature that varying spatial resolution of a single modeling project in the same study area may cause direct impact upon model predictions for flow and water quality outputs (Chaubey et al., 2005). In this study, it was further investigated if the modeling results could be inconsistently affected by alternative watershed simulation models even initialized by the same data resolution. Three watershed models were used to evaluate and compare the impacts of three alternative future land development scenarios for Queenstown, MD; a small (37 km^2) coastal community located on the Chesapeake Bay's Eastern Shore (Figure 1). The models were the Soil and Water Assessment Tool (SWAT) (Arnold et al., 2012), the Generalized Watershed Loading Function (GWLF) model (Haith and Shoemaker, 1987) and the Chesapeake Bay Program's Chesapeake Watershed Model (CBP-CWM) (Linker et al., 2013). It was stated in literature that the SWAT model is slightly better than GWLF in terms of nutrient predictions. However, both models performed similarly in hydrological processes (Niraula et al., 2013). In this study, model predictions of flow, total nitrogen (TN), and total phosphorus (TP) under different LULC configurations were compared; and model predictions were combined into ensemble averages, which were also compared to the predictions of the individual models.

2. Materials and Methods

2.1. Study Area

Queenstown is located within the Chesapeake Bay drainage, in Coastal Plain physiographic province of Maryland (Figure 1). The study site has relatively flat terrain with elevations ranging from 0 to 10 meters above mean sea level (AMSL). Because of the affordable land, low taxes, and proximity to the Washington DC and Baltimore metropolitan areas; the area's population is likely to increase by as much as 50 percent over the next 20 to 30 years (Jantz et al., 2010).

The study area consists of three watersheds (Figure 1). The Queenstown Harbor Links watershed is the smallest (4.7 km^2), including only small 0 or 1 order concentrated flow delivered to the Chester River subestuary. Land use includes a golf course and resort and overall land cover is 44% open urban land, 24% croplands and 24% forests (Table 1) (2002 Maryland LULC inventory http://planning.maryland.gov/OurWork/landuse.shtml). The Queenstown Creek watershed (QT Creek) is mainly croplands (57%) and forests (23%) with a relatively small percentage of developed lands (14%). The Upper Wye watershed is the largest watershed (24 km^2) in the study area and has 60% cropland, 25% forest; and 10% developed land. The Queenstown Harbor Links and QT Creek watersheds drain directly to the Chester River subestuary, while the Upper Wye watershed drains south through the Wye River to the Eastern Bay subestuary. The current Queenstown municipality is in the QT Creek watershed. Planning scenarios direct the bulk of development to the QT Creek and Upper Wye watersheds with no further development in the Queenstown Harbor Link watershed.

2.2. LULC Scenarios

To assess potential impacts from future development, a baseline scenario representing current conditions and three alternative future LULC scenarios were developed with the Queenstown Planning Commission (Table 1 and Figure 2). The “Distributed Growth” scenario (DG) assumes low intensity urban development across the entire planning area to the levels permitted by the current county zoning. Housing density would range between 1 and 20 acres per residential unit. In the two “Consolidated Growth” scenarios (High Impact Consolidated Growth (HI-CG) and Low Impact Consolidated Growth (LO-CG)), development would occur in designated areas adjacent to the current municipality while outlying areas would remain cropland, pasture or forest. The consolidated build-out capacity was defined by assuming a 50 percent increase in development over the county zoning with additional commercial development to support the residential growth. The designated growth area was defined by assuming a housing density similar to the existing municipality targeted to locations adjacent to the current municipality but not in sensitive or flood-prone areas, such as wetlands and areas within 300 meters of a stream. Approximately 70 percent of the Queenstown planning area would remain open space in the consolidated growth scenarios. The HI-CG and LO-CG scenarios differ in the land management of that open space. In the HI-CG scenario, the open space would be used for row crops whereas in LO-CG scenario, the open space would be used as pasture.

2.3. Watershed Models

This section provides an overview and further references to the three watershed models used to evaluate the impacts of alternative scenarios on water and nutrient discharges.

2.3.1 Soil and Water Assessment Tool (SWAT)

SWAT is a semi-distributed, process-based hydrologic and water quality model (Arnold et al., 2012) developed by U.S. Department of Agriculture (USDA). SWAT can assess long term impacts of management practices and climate change scenarios in complex watersheds. Major model components in SWAT include hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides, and agricultural management (Borah and Bera, 2003; Niraula et al., 2012; Wang et al., 2013, 2016; Yen et al., 2016).

In SWAT, hydrologic processes are simulated daily for hydrologic response units (HRU), which are areas with similar LULC, management, and soil attributes that are distinct from other HRUs. Runoff volume is simulated using the Soil Conservation Service's Curve Number Method (Mockus, 1969) or the Green and Ampt infiltration equation (1911). Potential evapotranspiration (PET) for each HRU can be estimated from soil permeability and vegetation cover using three different methods, and then adjusted into actual evapotranspiration based on expected soil moisture content. Empirical equations are utilized for modeling groundwater flow. Sediment yield is computed using the MUSLE equation (Modified Universal Soil Loss Equation, Williams and Berndt, 1977). SWAT models nitrogen using three organic pools (residue, stable, and active nitrogen) and two inorganic pools (ammonia and nitrate). Mineralization, nitrification, denitrification, and volatilization govern the balance among the different pools. The nitrate concentrations in runoff, lateral flow, and percolation are functions of the volume of water and the average concentration of nitrate in the soil layer (Neitsch et al., 2005). Phosphorus is divided into two organic pools (fresh residue and humic substance) and three mineral pools (labile in solution, labile on the soil surface and fixed in soil) with decay and mineralization moving P among the pools. The soluble P concentration in surface runoff is a function of the labile P concentration in the top soil layer, runoff volume, and a partitioning factor. Concentrations of sediment-bound N and P are functions of sediment yield and organic nutrient concentration in top soil layer. SWAT has been widely applied across many disciplines with over 2000 peer reviewed publications (SWAT Literature Database, 2015), including applications in US Mid-Atlantic and Northeast regions (Meng et al. (2010); Chu et al. (2004).

2.3.2 Generalized Watershed Loading Function (GWLF)

The GWLF model (Haith and Shoemaker, 1987) simulates runoff and sediment delivery using the Curve Number method and the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978). Nutrient loads are estimated from export coefficients for different LULC. GWLF also has algorithms for calculating septic system loads and for including point source discharge data. The model uses daily time steps for weather data and water balance calculations and produces monthly discharge and nutrient loads by aggregating daily model estimates into monthly values. Site-specific calibration is achieved by adjusting the parameters that control flow separation between storm flow and base flow, deep seepage, nutrient transport, soil erosion, and sediment delivery. GWLF is suitable for estimating source loads and total discharges at seasonal and inter-annual time-scales, and it has been used in TMDL development nationally (Borah et al., 2006) and in the mid-Atlantic and northeast regions (Fisher et al., 2006; Li et al. 2009).

2.3.3 CBP-CWM

The Chesapeake Bay Program's Chesapeake Watershed Model (CBP-CWM) is the regulatory model used to develop the Chesapeake Bay TMDL allocations and to assess which alternate scenarios of LULC and land management practices can best meet nutrient and sediment reduction goals.

The hydrological component of the CBP-CWM is based on the HSPF model (Hydrologic Simulation Program FORTRAN, Bicknell (2001)), which is a lumped parameter simulation model that uses hourly meteorological data to drive water transport and storage through watershed segments. Modeled components include surface-water runoff, surface depression storage, ground-water flow, evapotranspiration, and interception by vegetation. Landscape attributes, including topography, LULC and hydrography are used to define hydrologic response parameters that control the partitioning of water among different flow routes. Nutrient and sediment loads from major LULCs and the load responses to management practices are simulated with integrated sub-models. Organic and inorganic N cycles are simulated with a mechanistic model of the biogeochemical processes that regulate the transfer of land surface N additions to different soil, water, and atmospheric pools. P constituents are modeled using export coefficients that vary with LULC and soil properties and are applied to calculate the percentage of the P pool that is delivered to streams. Total phosphorus (TP) delivery is closely associated with sediment delivery, which is estimated from USLE erosion rates (Linker et al., 2013). For the model evaluation, CBP-CWM estimated discharges were compared to GWLF and SWAT predictions directly. For the Queenstown planning scenario assessment, CBP-CWM predicted loading rates for the relevant land-river segments were applied by LULC class across the Queenstown HUCs (see Boomer et al., 2011 for more details).

2.4. Model Setup, Calibration, and Validation

The watershed models were driven with inputs from meteorological, topographic, LULC, and soils datasets. Hourly and daily weather data for 1984-2005 (precipitation, temperature, wind speed, relative humidity, dew point temperature, solar radiation, and cloud cover) were acquired from the Chesapeake Bay Environmental Observatory's database (CBEO, 2012). A 10-meter DEM of the region was extracted from the USGS National Elevation Database (http://ned.usgs.gov/) and used to derive topographic inputs. Soil properties were obtained from the Soil Survey Geographic database (SSURGO) for Queen Anne's County, MD (http://soils.usda.gov/survey/geography/ssurgo). Current LULC came from the 2002 Maryland LULC inventory supplied state of Maryland Department of Planning (http://planning.maryland.gov/OurWork/landuse.shtml).

Flow and water quality data were not available for the Queenstown study watersheds, so the three models were calibrated and validated with measured flow, TN, and TP discharges from six gauged watersheds located approximately 20 km east of the study area (Figure 3). These six watersheds (304, 305, 306, 310, Greensboro, and Ruthsburg) were monitored for flow and water quality (TN and TP) for multiple years between 1984 and 2005, and collectively they provide over 30 years of flow and water quality data (Jordan et al. 1997; http://cbrim.er.usgs.gov/). Three watersheds (304, 310 and Greensboro) were used to calibrate the models, and the other three watersheds (305, 306, and Ruthsburg) were used to validate the models. Calibration and validation were performed at the monthly timescale. Essential characteristics such as average elevation, average slope, and hydrologic soil groups of the targeted watershed are shown in Table 2.

2.5. Model Comparisons and Synthesis

The calibrated and validated models were applied to the Queenstown study area to quantify the effects of current LULC and of the three future land management scenarios on flow and water quality outputs. The predictions of the models were combined into ensemble predictions using weighted averaging (see below), and the ensemble predictions of the scenarios were compared to identify the least detrimental future LULC scenario. The weights were assigned based on the model performance at the validation sites. Concordance among the three models was measured with a variation index that was estimated separately for each constituent (flow, N, or P) at each time step (month or year):

$$Variation\ index\ (\vartheta) = \frac{1}{n} \times \sum_{1}^{n} |X_i - \bar{X}| \tag{1}$$

where, i is the model index, n is the number of predictions (models) available for the constituent at a specific time; $\bar{X}$ is the average of those n predictions, and X_i is the i^{th} prediction. Small values of ν indicate close agreement among model outputs and large values indicate disagreement.

Models were assigned weights for each constituent based on performance at the validation sites (305, 306 and Ruthsburg, Figure 3), such that:

$$\lambda_{i,j} = \frac{e^{\left(Ens_{i,j}-1\right)}}{\sum_{i=1}^{n} e^{\left(Ens_{i,j}-1\right)}} \tag{2}$$

where $\lambda_{i,j}$ is the weight assigned to model *i* for constituent *j*, *Ens* is the Nash–Sutcliffe efficiency from model validation, and *n* is the number of models (3). *Ens* can theoretically range from $-\infty$ to 1. Values near 1 indicate near perfect agreement between model predictions and observed data, values near 0 indicate that the model is no better than simply using the average of the data, and negative values indicate that the model is worse than using that average. For a given constituent *j,* the weights $\lambda_{i,j}$ sum to 1. Single model predictions for the Queenstown assessment area were combined into ensemble predictions for each constituent and each scenario using Eq. (2), and those model average outputs were used to identify the least detrimental LULC Scenarios.

3. Results and Discussions

3.1. General Statistics for Model Calibration and Validation

In Table 3, goodness of fit results (R^2 and *Ens*) are presented in calibration and validation sites for all models. In addition, time series of observed data compared with model predicted flow, TP, and TN fluxes at calibration and validation watersheds are presented in the Appendix (Figure A1~A6). All models performed well in predicting flow, with average *Ens* values around 0.7 and 0.6 at the calibration and validation sites. Nitrogen predictions also had good but slightly lower *Ens* values (~0.6 and 0.5 for calibration and validation sites, respectively, Table 3). For phosphorus, the models had some negative average *Ens* values at the calibration sites (mostly at site 310), but the performance was acceptable at the validation sites (average *Ens*=0.2, Table 3). All three models are best at predicting flow (high *Ens*), intermediate at predicting TN (moderate

Ens), and poor at predicting TP (low or negative *Ens*, Table 3). In addition to R^2 and *Ens*, mass balance error (MBE) was also tested to examine the potential differences among statistical measures. As shown in Table 3, coherent responses of MBE can be found in comparing with two other statistics. In general, it is hard to single out a specific model with better or poor performance in terms of statistical results.

3.2. Simulation Results with Current LULC Map

According to the variation index (Eq. 1), flow and TN predictions for current conditions (1984-2005) were less variable among models than were TP predictions (Figure 4). Except for the first year (1984), the variation index values for flow predictions were less than 0.25 and those for TN were less than 0.4, while index values for TP were higher (up to 0.74). As expected, all of the models predicted higher discharge during wet years (e.g., 1989, 1996, 1999 and 2003) than in drier years, but there also was greater variation among model predictions in wetter years. SWAT and GWLF had the highest and lowest predictions, respectively, for TP among the three models. For flow, SWAT and CBP-CWM predicted higher mean annual discharge (45 to 50 cm/year) than GWLF (32 cm/year).

The SWAT, GWLF and CBP-CWM models follow similar temporal patterns in monthly predictions (Fig. 5). Flow is maximum around March and minimum in August. Flow predictions are most consistent among models in the wetter winter and spring months (December – May). The highest variation in predicted flow among models occurs in the summer to early fall (July-September). In August, GWLF’s flow prediction is about one fourth of the SWAT and CBP-CWM predictions. TN and TP predictions follow similar monthly patterns. Variation among the model predictions is lower in winter and spring compared to summer and fall, and the highest variation occurs in July and August, the driest months of the year. The variation index is notably higher for TP than for TN and flow, due to relatively the large difference between SWAT and GWLF predictions.

The patterns of variation among model predictions in a wet year (2003) are different from the patterns in a dry year (1987, Fig. 6). For the dry year (1987, 90 cm of precipitation), variations in flow predictions were low (less than 0.20) in the winter months, and substantially higher in the dry months (July through November). January has the highest average predicted flow among all months and the smallest variation among the models. SWAT predicts a February high peak flow, which may indicate that SWAT is relatively more sensitive to seasonal events (snow melt in this case) and the potential corresponding groundwater contribution. CBP-CWM predicts higher TN fluxes during a dry year than either the SWAT or GWLF models. For TP, the GWLF and CBP-CWM models predict similarly low loads that vary with the flow pattern, whereas SWAT oscillates significantly over the year with four local peaks. GWLF and CBP-CWM predict extremely low TP loads from March to December (spring, summer, and fall). In addition, January and February have distinctively higher TP loads.

For the wet year (2003, 168 cm of precipitation), the variation in flow predictions is generally low, and the highest variation occurs during February and the summer months when SWAT predicts higher discharge. Regarding model simulations in TN, both SWAT and GWLF predicted temporal patterns of TN loads similar to the patterns of flow simulation. CBP-CWM attributes almost all of the TN loads to groundwater delivery (baseflow), and therefore predictions fluctuate only marginally over the year. For TP, the pattern of monthly discharge in the wet year is similar to average monthly TP discharge. SWAT has a large peak in February, when GWLF has a smaller peak. The high TP and TN peaks result from higher predicted flows in February, but may also reflect fertilizer applications during that month (Zhu et al., 2012). The

GWLF and CBP-CWM models do not explicitly account for monthly variation in fertilizer application.

3.3. LULC Scenario Analysis

3.3.1 Annual Predictions of Hydrological & Nutrient Processes

The differences among LULC scenario predictions for any model were relatively small compared to the differences among models for any LULC scenario (Figure 7 and 8). The predicted impacts of development on flow and nutrients delivered to the Queenstown Harbor Links watershed and Upper Wye River were similar. A common approach of scenario analysis is to look at the change of flow and nutrient loadings relative to a baseline scenario (Huisman et al., 2009). In this study, the current LULC scenario is the baseline scenario, and all the changes were calculated relative to that baseline (Fig. 8). Changes in LULC in the Queenstown Harbor Links watershed are not expected (Table 1), so relative changes in flow, TN and TP were not assessed for this area. We expected similar directions of response to the LULC changes among all three models, but likely different rates or magnitude of response. The responses were more complicated than we expected, and in some cases there are almost no changes in discharges or loadings despite shifts in LULC conditions. The trends in predictions are interpreted separately for flow, TN and TP.

Flow: SWAT predicted that development would increase stream discharge by as much as 6 to 9%, and that distributed growth would have the greatest impact on average annual flow (Figs. 7 and 8). In contrast, CBP-CWM predicted that any future development would decrease annual average discharge by as much as 3%, with the consolidated growth scenarios having the biggest impact. GWLF flow predictions varied less than 1% across all scenarios. It has been shown previously that SWAT may generate higher peak flow during the winter/spring seasons (due to potential snow melt events). However, this issue can also be justified in literature since urbanization is known to have the corresponding increase of flow (Owe, 1985).

Nitrogen: For all three LULC scenarios, SWAT and GWLF predict TN increases up to 6%, while CBP-CWM predicted TN decreases of 7.5% for the "HI-CG" scenario and 17% for the "LO-CG" scenario. Overall, SWAT and GWLF tend to agree on both the direction and magnitude of TN change (except for QT Creek watershed). CBP-CWM predicted a decrease in TN for all scenarios in all watersheds. For TN loads, "LO-CG" was predicted by SWAT to be the least environmental friendly development scenario, but was the most environmental friendly according to CBP-CWM GWLF, which predicted that "HI-CG" was the least favorable scenario.

Phosphorus: The highest agreements among the three watershed models are observed in relative changes in phosphorus prediction in Queenstown, but agreement was not as good in the Upper Wye and QT Creek watersheds. Almost all three models predict lower TP loadings for future scenarios at the whole study area (except for one in Upper Wye). SWAT predicts up to 10% higher TP loading for QT creek, whereas the other two models report TP reduction.

3.3.2 Least Detrimental LULC Scenarios

Weights (λ) assigned to each model (for each constituent) based on their performance at the validation sites are shown in Table 4. Once the three models predictions on current and future LULC scenarios were synthesized by the method presented earlier, the relative changes in water quality and quantity caused by converting the current LULC to each of future LULC scenario were calculated (Table 5). The environmental impacts of the three development scenarios were ranked using ensemble averages of the predictions from the three models, where the models were weighted by their performance in model validation (Table 4). The Distributed Growth (DG) scenario will reduce the TN and TP by 2.8% and 7.2%, respectively (Table 5), and appears to be the development scenario with relatively better performance (i. e., it has the lowest nutrient loads). On the other hand, the DG scenario is closely followed by LO-CG scenario with 2.2% and 7.8% reductions of TN and TP (Table 5). DG has the highest reduction for TN, but LO-CG has the highest reduction for TP indicated the fact that the complexity of three implemented watershed models and resolution of our understanding currently are not yet suited to provide reliable suggestion for the following acts as a part of the decision making processes (e.g., law making, environmental protection regulations, or conservation practices).

4. Summary and Conclusions

In this study, three watershed models were applied to Queenstown, MD (a coastal community on Maryland's Eastern Shore of Chesapeake Bay) to evaluate the potential impacts of anthropogenic development on flow, TN and TP loadings to the Chesapeake Bay. Three models performed similarly during calibration and validation among LULC scenarios. However, it is hard to identify which model may provide consistently better results (model predictions in terms of statistics) than the other. Similar findings also have been reported by Niraula et al. (2013) when comparing SWAT with GWLF, whereas neither of the models was significantly better than the other in simulating flow, sediment and nutrient loads.

In general, there was a good agreement on annual average flow for Queenstown between the SWAT and CBP-CWM models; GWLF and CBP-CWM predicted similar TN and TP loads. Each model has different strengths and weaknesses. For instance, the primary strength of the SWAT model is that SWAT has numerous empirically- and physically based functions that govern complex hydrologic and nutrient processes. SWAT is capable of simulating the targeted watershed with proper settings. In addition, it has more than 2,000 peer-reviewed journal articles supported as solid information base. It is fairly easy to solve challenging tasks within short timeframe. However, it could also be the weakness since it requires large number of system parameters. Users may face challenging calibration issues such as high-dimensional problems and it may be over-calibrated in some cases. On the other hand, GWLF is the model among the three that requires the least information from users. The associated benefits and drawbacks are right exactly the other way of SWAT. The CBP-CWM model, which is based upon the HSPF, is right in between SWAT and GWLF which compensate the computational loads from system parameters with modeling performance in terms of simulation precision. Therefore, model predictions were combined into an ensemble prediction weighted by model performance at the validation sites. It was stated in literature that major sources of uncertainty in watershed modeling are forcing inputs, system parameters, measurement data, and model structure (Yen et al., 2014a). The implementation of applying combinations of LULC with different models is also the exploration of structural uncertainty. In this study, both structural and input uncertainty was incorporated to examine the potential impacts upon model predictions. Using a combination of LULC allowed us to understand the relative importance of different hydrologic processes among the models (and accordingly, major sources of uncertainty).

The use of multiple models and combining outputs in a systematic manner is gaining wider acceptance (Yen et al., 2015). For example, the Western Lake Erie Basin has been investigated by five research groups to explore higher level of scientifically credible and practice solutions for upcoming environmental issues (Scavia et al., 2016). This study demonstrated the benefits of using multiple models to assess the potential impacts of LULC change and the corresponding concurrent impacts on flow and nutrient processes. The use of multiple models or model ensembles may significantly improve the reliability on predictions and could/should be extended to programs like TMDL development and NPDES permitting.

References

Agardy, T. and Alder, J. (2005). "Coastal systems." Millennium Ecosystem Assessment Series, 1.

Amiri, B. J. and Nakane, K. (2009). "Comparative prediction of stream water total nitrogen from land cover using Artificial Neural Network and multiple linear regression approaches." Polish Journal of Environmental Studies, 18, 151-160.

Arnold, J., D. Moriasi, P. Gassman, K. Abbaspour, M. J. White, R. Srinivasan, C. Santhi, R. D. Harmel, A. van Griensven A., M. W. van Liew, N. Kannan and M. K. Jha, (2012). SWAT: Model use, calibration, and validation. Transactions of the ASABE. 55(4): 1491-1508.

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C. and Silliman, B. R. (2011). "The value of estuarine and coastal ecosystem services." Ecological Monographs, 81(2), 169-193.

Beven, K. J. and Freer, J. (2001). "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology." Journal of Hydrology, 249, 11-29.

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Donigian, A. S. and Johanson, R. C. (2001). "Hydrologic 899 Simulation Program - FORTRAN (HSPF): User's Manual for Release 12.", Aqua Terra 900 Consultants, Mountain View, CA.

Boesch D.F., Brinsfield, R.B., Magnien, R.E. (2001). “Chesapeake Bay eutrophication: scientific understanding, ecosystem restoration, and challenges for agriculture.” J Environ Qual. 30(2):303-20.

Boomer, K., Li, X., Weller, D. E., and Kalin, L. 2011. “Applying Multiple Models to Inform Watershed Management and Land Use Planning in a Coastal Community.” Final Report Submitted to The NOAA/UNH Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET).

Boomer, K., Weller, D. E., Jordan, T. E., Linker, L., Liu, Z.-J., Reilly, J., Shenk, G. and Voinov, A. A. (2013). “Using Multiple Watershed Models to Predict Water, Nitrogen, and Phosphorus Discharges to the Patuxent Estuary.” JAWRA Journal of the American Water Resources Association, 49: 15–39. doi: 10.1111/j.1752-1688.2012.00689.x

Borah, D., Yagow, G., Saleh, A., Barnes, P., Rosenthal, W., Krug, E. and Hauck, L. (2006). "Sediment and nutrient modeling for TMDL development and implementation." Transactions of the ASAE, 49(4), 967-986.

Borah, D. K. and Bera, M. (2003). "Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases." Transactions of ASAE, 46(6), 1553-1566.

Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M. and Viney, N. R. (2009). "Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use." Advances in Water Resources, 32(2), 129-146.

Chaubey, I., Cotter, A. S., Costello, T. A. and Soerens, T. S., (2005), Effect of DEM data resolution on SWAT output uncertainty. Hydrol. Process., 19: 621–628. doi: 10.1002/hyp.5607

Chesapeake Bay Program (2010). "Bay Barometer; A Health and Restoration Assessment of the Chesapeake Bay and Watershed in 2009."

Chu, T., Shirmohammadi, A., Montas, H. and Sadeghi, A. (2004). "Evaluation of the SWAT model's sediment and nutrient components in the Piedmont physiographic region of Maryland."

Exbrayat, J. F., Viney, N R., Seibert J, Wrede S, Frede, H. G. and Breuer, L. (2010). "Ensemble modelling of nitrogen fluxes: data fusion for a Swedish meso-scale catchment." Hydrology and Earth System Sciences, Discuss., 7, 5299-5334.

Fisher, T. R.,D, H. J., Boynton, W. R. and Williams, M. R. (2006). "Cultural eutrophication in the Choptank and Patuxent estuaries of Chesapeake Bay." Limnology and Oceanography, 51, 435-447.

Georgakakos, K. P., Seo, D. J., Gupta, H., Schake, J. and Butts, M. B. (2004). "Characterizing streamflow simulation uncertainty through multimodel ensembles." Journal of Hydrology, 298, 222-241.

Goswami, M., O'Connor, K. M., Bhattarai, K P. and Shamseldin, A Y. (2005). "Assessing the performance of eight real-time updating models and procedures for the Brosna River." Hydrology and Earth System Sciences, 9, 394-411.

Green, W. H. and Ampt, G. A. (1911). "Studies on soil physics, 1. The flow of air and water through soils." Journal of Agricultural Sciences, 4, 11-24.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J. G., Bai, X. M. and Briggs, J. M. (2008). "Global change and the ecology of cities." Science, 319, 756-760.

Grizzettia, B., Bouraoui, F. and De Marsily, G. (2005). "Modelling nitrogen pressure in river basins: A comparison between a statistical approach and the physically-based SWAT model." Physics and Chemistry of the Earth, 30(8-10), 508-517.

Haith, D. A. and Shoemaker, L. L. (1987). "Generalized Watershed Loading Functions for Stream Flow Nutrients." Water Resources Bulletin, 23(3), 471-478.

Huisman, J. A., Breuer, L., Bormann, H., Bronstert, A., Croke, B. F., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M., Viney, N. R. and Willems, P. (2009). "Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis." Advances in Water Resources, 32(2), 159-170.

Jantz, C., Goetz, S., Donato, D. and Claggett, P. (2010). "Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model." Computers, Environment and Urban Systems(34), 1-16.

Jordan, T. E., D. L. Correll, and D. E. Weller. 1997. Effects of agriculture on discharges of nutrients from coastal plain watersheds of Chesapeake Bay. J. Environ. Qual. 26:836-848.

Kadane, J. B. and N. A. Lazar, (2004). Methods and Criteria for Model Selection. Journal of the American Statistical Association 99:279-290.

Kalin, L. and Hantush, M. H., (2006). "Comparative Assessment of two distributed watershed models with application to a small watershed." Hydrological Processes, 20(11), 2285-2307.

Kemp, W., Boynton, W., Adolf, J., Boesch, D., Boicourt, W., Brush, G., Cornwell, J., Fisher, T., Gilbert, P., Hagy, J., Harding, L., Houde, E., Kimmel, D., Miller, W., Newell, R., Roman, M., Smith, E. and Stevenson, J. (2005). "Eutrophication of Chesapeake Bay: historical trends and ecological interactions." Marine Ecology Progress Series, 303, 1-29.

Li, Z., Xu, Z., Shao, Q. and Yang, J. (2009), Parameter estimation and uncertainty analysis of SWAT model in upper reaches of the Heihe river basin. Hydrol. Process., 23: 2744–2753. doi: 10.1002/hyp.7371

Linker, Lewis C., Richard A. Batiuk, Gary W. Shenk, and Carl F. Cerco. 2013. Development of the Chesapeake Bay Watershed Total Maximum Daily Load Allocation. Journal of the American Water Resources Association (JAWRA) 49(5): 986–1006. DOI: 10.1111/jawr.12105

Maryland Department of Planning (2007). "Water Resources Element of the Comprehensive Plan Guidance Document Planning for Water Supply, Wastewater Management and Stormwater Management." Maryland Department of Planning, Baltimore, MD, 84.

Maryland Office of Planning (1993). "Nonpoint source assessment and accounting system: Final report for the FFY '91 Section 319 grant. ." Maryland Office of Planning, Baltimore, MD.

Maryland Office of Planning (1995). "Development and application of the nonpoint source assessment and accounting system: Final report for the FFY '92 Section 319 grant.", Maryland Office of Planning,, Baltimore, MD.

Meng, H., Sexton, A., Maddox, M., Sood, A., Brown, C., Ferrara, R. and Murtugudde, R. (2010). "Modeling Rappahannock River Basin Using SWAT—Pilot for Chesapeake Bay Watershed." Applied engineering in agriculture, 26(5), 795.

Mockus, V. (1969). "Hydrologic soil-cover complexes." SCS National Engineering Handbook: Section 4. Hydrology, 10.11-10.24.

Nagy, R. C., Lockaby, B. G., Kalin, L. and Anderson, C. (2012). "Effects of urbanization on stream hydrology and water quality: the Florida Gulf Coast." Hydrological Processes.

Nash, J. E. and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models. Part I- a discussion of principles." Journal of Hydrology, 10, 282-290.

Nasr, A., Bruen, M., Jordan, P., Moles, R., Kiely, G. and Byrne, P. (2007). "A comparison of SWAT, HSPF, and SHETRAN/GOPC for modeling phosphorus export from three catchments in Ireland." Water Research, 41(5), 1065-1073.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, R. and Williams, J. R. (2005). Soil and Water Assessment Tool, Theoretical Documentation: Version 2005, USDA Agricultural Research Service and Texas A&M Blackland Research Center, Temple, Texas.

Niraula, R., L. Kalin, R. Wang, and P. Srivastava (2012), " Determining Nutrient and Sediment Critical Source Areas with SWAT Model: Effect of Lumped Calibration", Transactions of ASABE, 55(1):137-47

Niraula, R., K. Latif, P. Srivastava, C.J. (2013). Anderson Identifying critical source areas of nonpoint source pollution with SWAT and GWLF Ecol. Model., 268, pp. 123–133.

Own, M. (1985). "Long-term streamflow observations in relation to basin development." Journal of Hydrology, 78(3-4), 243-260.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. J. and al., e. (2004). "Overall distributed model intercomparison project results." Journal of Hydrology, 298, 27-60.

Renner, M., Werner, M. G., Rademacher, S. and Sprokkereef, E. (2009). "Verification of ensemble flow forecast for the River Rhine." Journal of Hydrology, 376, 463-475.

Rose, S. (2002). "Comparative major ion geochemistry of Piedmont streams in the Atlanta, Georgia region: Possible effects of urbanization." Environmental Geology, 42, 102-113.

Scavia, D., M. Kalcic, R. L. Muenich, J. Read, N. Aloysius, C. Boles, R. Confessor, J. DePinto, M. Gildow, J. Martin, T. Redder, D. Robertson, S. Sowa, Y. Wang, H. Yen. 2016. Shaping Lake Erie agriculture nutrient management via scenario development. University of Michigan, Ann Arbor. http://tinyurl.com/pp4umuz

Schoonover, J. E., Lockaby, B. G. and Pan, S. (2005). "Changes in chemical and physical properties of stream water across an urban-rural gradient in western Georgia." Urban Ecosystem, 8, 107-124.

Sharifi, A., M. Lang, G. W. McCarty, A. M. Sadeghi, S. Lee, H. Yen, M. C. Rabenhorst, J. Jeong, I. Yeo, (2016) “Improving Model Prediction Reliability through Enhanced Representation of Wetland Soil Processes and Constrained Model Auto Calibration - A Paired Watershed Study” Journal of Hydrology , 541, Part B, pp. 1088-1103. DOI: 10.1016/j.jhydrol.2016.08.022

Shen, Z. Y., Gong, Y. W., Li, Y. H., Hong, Q., Xu, L. and Liu, R. M. (2009). "A comparison of WEPP and SWAT for modeling erosion of the Zhangjiachong watershed in the Three Gorges Reservoir Area." Agricultural Water Management, 96, 1435–1442.

SWAT Literature Database (2015). "SWAT Literature Database for Peer-Reviewed Journal Articles." <https://www.card.iastate.edu/swat_articles/>. (November, 2012).

Tu, J. (2009). "Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA." Journal of Hydrology, 379(3-4), 268-283.

USDA, S. (1986). "Urban hydrology for small watersheds." Technical Release, 55, 2-6.

Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F., Frede, H., Gräff, T., Hubrechts, L., Huisman, J. A., Jakeman, A. J., Kite, G. W., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M. and Willems, P. (2009). "Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions." Advances in Water Resources, 32(2), 147-158.

Wahl, M. H., McKellar, H. N. and Williams, T. M. (1997). "Patterns of nutrient loading in forested and urbanized coastal streams." Journal of Experimental Marine Biology and Ecology, 213(1), 111-131.

Wang, R., L. Kalin, W. Kuang, H. Tian. (2013). Individual and Combined Effects of Land Use/Cover and Climate Change on Wolf Bay Watershed Streamflow in Southern Alabama. Hydrological Processes, DOI: 10.10102/hyp.10057

Wang, R., L.C. Bowling, and K.A. Cherkauer (2016). Estimation of the Effects of climate variability on Crop yield in the Midwest USA. Agricultural and Forest Meteorology, 216, 141–156. doi:10.1016/j.agrformet.2015.10.001.

Weller, D. E., and M. E. Baker. (2014). “Cropland riparian buffers throughout Chesapeake Bay watershed: spatial patterns and effects on nitrate loads delivered to streams.” J. Am. Water Res. Assn. 50:696-715.

Weller, D. E., M. E. Baker, and T. E. Jordan. (2011). Effects of riparian buffers on nitrate concentrations in watershed discharges: new models and management implications.” Ecol. Applic. 21:1679-1695.

Williams, J. R. and Berndt, H. D. (1977). "Sediment yield prediction based on watershed hydrology." Transactions of ASAE, 20(6), 1100-1104.

Winchell, M., Srinivasan, R., di Luzio, M. and Arnold, J. (2007). ArcSWAT interface for SWAT 2005. User’s Guide, Blackland Research Center, Texas Agricultural Experiment Station, Temple.

Wischmeier, W. H. and Smith, D. D. (1978). Predicting Rainfall Erosion Losses: A Guide to Conservation Planning, Agriculture Handbook No. 537. USDA/Science and Education Administration, US. Govt. Printing Office, Washington, DC.

Yen, H., Wang, X., Fontane, D. G., Harmel, R. D., Arabi, M. (2014a). A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling, Environmental Modelling and Software, 54, pp. 211-221, doi: 10.1016/j.envsoft.2014.01.004.

Yen, H., Bailey, R. T., Arabi, M., Ahmadi, M., White, M. J., Arnold, J. G. (2014b). The Role of Interior Watershed Processes in Improving Parameter Estimation and Performance of Watershed Models. Journal of Environmental Quality, published online. doi:10.2134/jeq2013.03.0110

Yen, H., A. Sharifi, L. Kalin, G. Mirhosseini, J. G. Arnold. (2015). Assessment of Model Predictions and Parameter Transferability by Alternative Land Use Data on Watershed Modeling. Journal of Hydrology, 527, pp. 458-470. DOI: 10.1016/j.jhydrol.2015.04.076

Yen, H., M. J. White, J. G. Arnold, S. C. Keitzer, M. V. Johnson, J. D. Atwood, P. Daggupati, M. E. Herbert, S. P. Sowa, S. A. Ludsin, D. M. Robertson, R. Srinivasan, C. A. Rewa (2016) “Western Lake Erie Basin: Soft-Data-Constrained, NHDPlus Resolution Watershed Modeling and Exploration of Applicable Conservation Scenarios.” Science of the Total Environment, 569-570, pp. 1265-1281. DOI: 10.1016/j.scitotenv.2016.06.202

Zhao, L., Duan, Q., Schaake, J., Ye, A. and Xia, J. (2011). "A hydrologic post-processor for ensemble streamflow predictions." Advances in Geosciences, 29, 51-59.

Zhu, Z., K. Broersma, A. Mazumder. (2012). “Impacts of Land Use, Fertilizer and Manure Application on the Stream Nutrient Loadings in the Salmon River Watershed, South-Central British Columbia, Canada.” Journal of Environmental Protection, 3, 809-822.

ACKNOWLEGMENTS

This project was partially funded by grants from (i) NOAA/UNH Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET); (ii) Center for Environmental Studies at the Urban Rural Interface; and (iii) United States Department of Agriculture - Natural Resources Conservation Service (USDA-NRCS) Conservation Effects Assessment Project (CEAP) - Wildlife and Cropland components. The calibration data were collected with support from NSF (BSR-89-05219, DEB-92-06811, and DEB-93-17968), the Governor's Research Council of Maryland, and the Smithsonian Institution Environmental Sciences Program.

Table 1. Land use percentages in the study watersheds‡

Scenario	Watershed	Area (km^2)	Land Use Type (%)				
			Urban	Forest	Cropland	Pasture	Other
Current	QT Harbor Link	4.7	31.4	35.8	23.9	0.0	8.9
	QT Creek	8.3	14.4	22.7	57.2	0.8	4.9
	Upper Wye	24	10.2	25.2	59.6	0.0	5.0
	Queenstown†	37	15.5	24.4	54.6	0.2	5.4
Distributed Growth (DG)	QT Harbor Link	4.7	31.4	35.8	23.9	0.0	8.9
	QT Creek	8.3	27.3	22.8	44.3	0.8	4.9
	Upper Wye	24	23.6	26.6	45.1	0.0	4.8
	Queenstown†	37	25.4	26.9	42.2	0.2	5.3
High Impact Consolidated Growth (HI-CG)	QT Harbor Link	4.7	31.4	35.8	23.9	0.0	8.9
	QT Creek	8.3	22.1	22.8	49.5	0.8	4.9
	Upper Wye	24	14.8	25.8	54.6	0.0	4.8
	Queenstown†	37	18.5	26.4	49.5	0.2	5.3
Low impact Consolidated Growth (LO-CG)	QT Harbor Link	4.7	31.4	35.8	23.9	0.0	8.9
	QT Creek	8.3	22.1	22.8	44.3	6.0	4.9
	Upper Wye	24	14.8	25.8	45.1	9.5	4.8
	Queenstown†	37	18.5	26.4	42.2	7.5	5.3

† The whole study area, consisting of the three watersheds altogether is referred to as "Queenstown"
‡ 2002 Maryland LULC inventory http://planning.maryland.gov/OurWork/landuse.shtml

Table 2. Essential characteristics of the targeted watersheds‡

Watershed	Area (km^2)	Average Elevation (m)	Average Slope (degree)	Hydrologic Soil Group (%)					
				A	B	B/D	C	C/D	D
304	10.2	19.2	1.9	0%	59%	7%	4%	23%	7%
305	17.8	18.2	1.8	0%	58%	10%	3%	23%	6%
306	7.5	21.6	1.2	5%	53%	10%	5%	23%	4%
310	54.7	19.1	1.5	1%	51%	4%	16%	19%	9%
Greensboro	294.0	17.3	1.5	2%	28%	0%	14%	1%	56%
Ruthsburg	59.0	60.3	1.42	14%	44%	0%	20%	4%	18%
Queenstown	37.1	9.7	2.2	0%	43%	1%	19%	31%	6%

† The whole study area, consisting of the three watersheds altogether is referred to as "Queenstown"
‡ 2002 Maryland LULC inventory http://planning.maryland.gov/OurWork/landuse.shtml

Table 3. Goodness of fit results at the calibration and validation sites

	Watershed &Time period	Variable	SWAT			GWLF			CBP-CWM		
			R^2	Ens	MBE[ψ]	R^2	Ens	MBE[ψ]	R^2	Ens	MBE[ψ]
Calibration Sites	304 Apr 89~Dec 92	Flow	0.70	0.67	-4%	0.74	0.77	-7%	0.67	0.66	-8%
		TN	0.58	0.36	-13%	0.65	0.56	9%	0.52	0.49	-10%
		TP	0.46	0.26	5%	0.28	-0.3	0%	0.31	0.10	-121%
	310 Jul 90~Oct 95	Flow	0.73	0.69	-8%	0.76	0.73	4%	0.77	0.75	-2%
		TN	0.74	0.54	5%	0.74	0.61	1%	0.80	0.77	1%
		TP	0.14	-1.66	6%	0.19	-0.92	15%	0.31	-0.07	-86%
	Greensboro Jan 84~Dec 99	Flow	0.70	0.67	-6%	0.74	0.70	0%	0.73	0.73	-2%
		TN	0.63	0.49	-2%	0.55	0.59	6%	0.75	0.73	-10%
		TP	0.29	0.11	0%	0.52	0.3	1%	0.49	0.45	-16%
Validation Sites	305 Apr 89~Dec 92	Flow	0.78	0.73	8%	0.62	0.6	-3%	0.65	0.64	-1%
		TN	0.50	0.44	-7%	0.65	0.50	6%	0.58	0.57	-8%
		TP	0.42	0.38	19%	0.41	-0.05	4%	0.28	0.18	-1%
	306 Apr 89~Feb 92	Flow	0.64	0.54	-21%	0.63	0.60	14%	0.38	0.34	-3%
		TN	0.65	0.59	10%	0.57	0.55	14%	0.34	0.34	-2%
		TP	0.62	0.21	8%	0.37	0.36	5%	0.16	0.10	-2%
	Ruthsburg † Nov 00~Mar 05	Flow	0.66	0.61	-9%	0.69	0.62	-1%	0.64	0.63	-9%

† Only flow data was available at this site
[ψ] Positive MBE (Mass balance error) indicates underestimation

Table 4. Weights (λ) assigned to each model and each constituent based on model performance at the validation sites. The weights were used to combine model predictions into an ensemble average.

	SWAT	GWLF	CBP-CWM
Flow	0.35	0.34	0.32
TN	0.34	0.34	0.32
TP	0.37	0.32	0.31

Table 5. Environmental impacts of the three land use scenarios estimated by the weighted ensemble average predictions. The numbers are the percentage change in discharge or load relative to the baseline current land use.

Variable	High Impact Consolidated Growth (HI-CG)	Low Impact Consolidated Growth (LO-CG)	Distributed Growth (DG)
Flow	0.36	0.21	2.22
TN	1.19	-2.17	-2.82
TP	-1.42	-7.81	-7.24

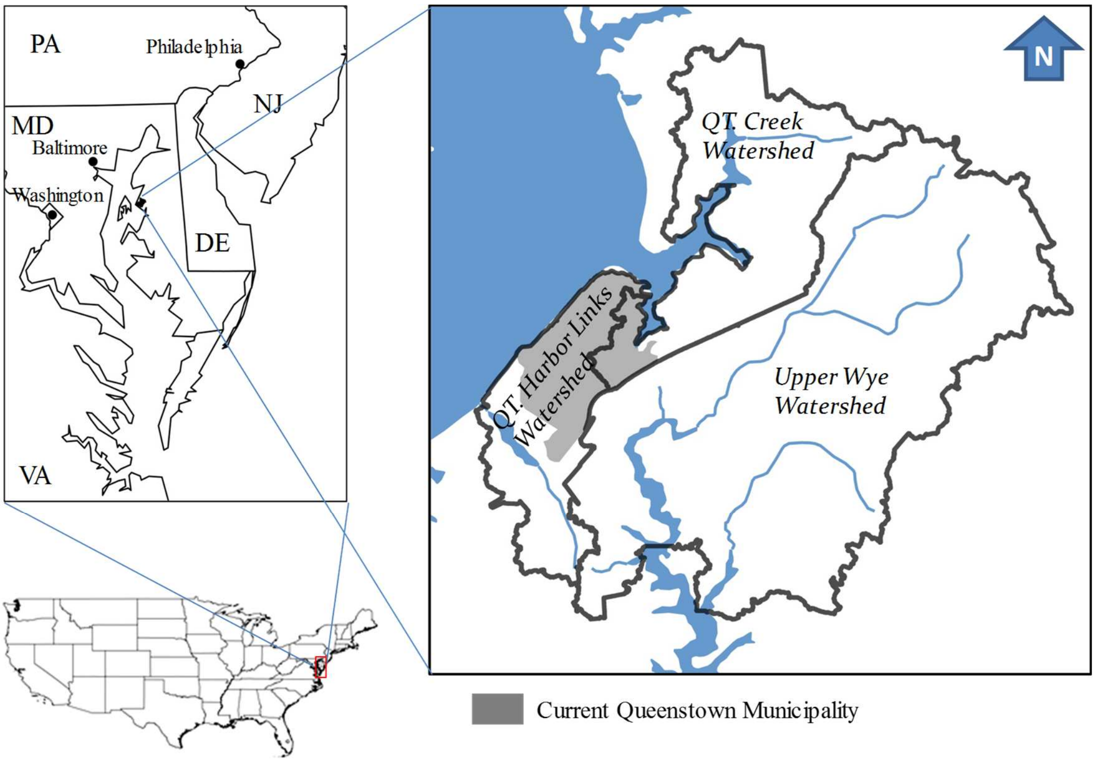

Figure 1. Three watersheds comprising the Queenstown study area on the eastern shore of Chesapeake Bay. Current development is mostly in the the gray area.

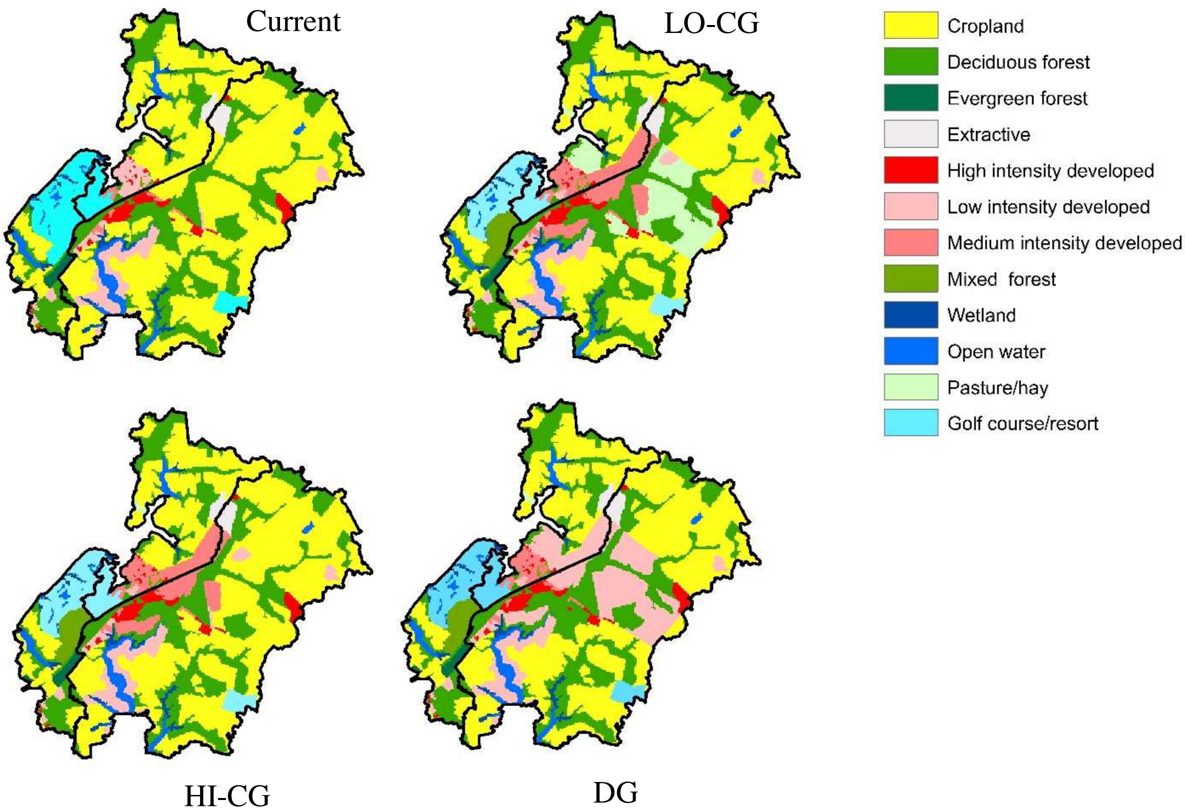

Figure 2. Land use maps for current conditions and for three future development scenarios. The distributed growth scenario allows low intensity development across a large area, whereas the two consolidated growth scenarios concentrate medium density development in a smaller area. DG: Distributed Growth; HI-CG: High Impact Consolidated Growth; LO-CG: Low Impact Consolidated Growth)

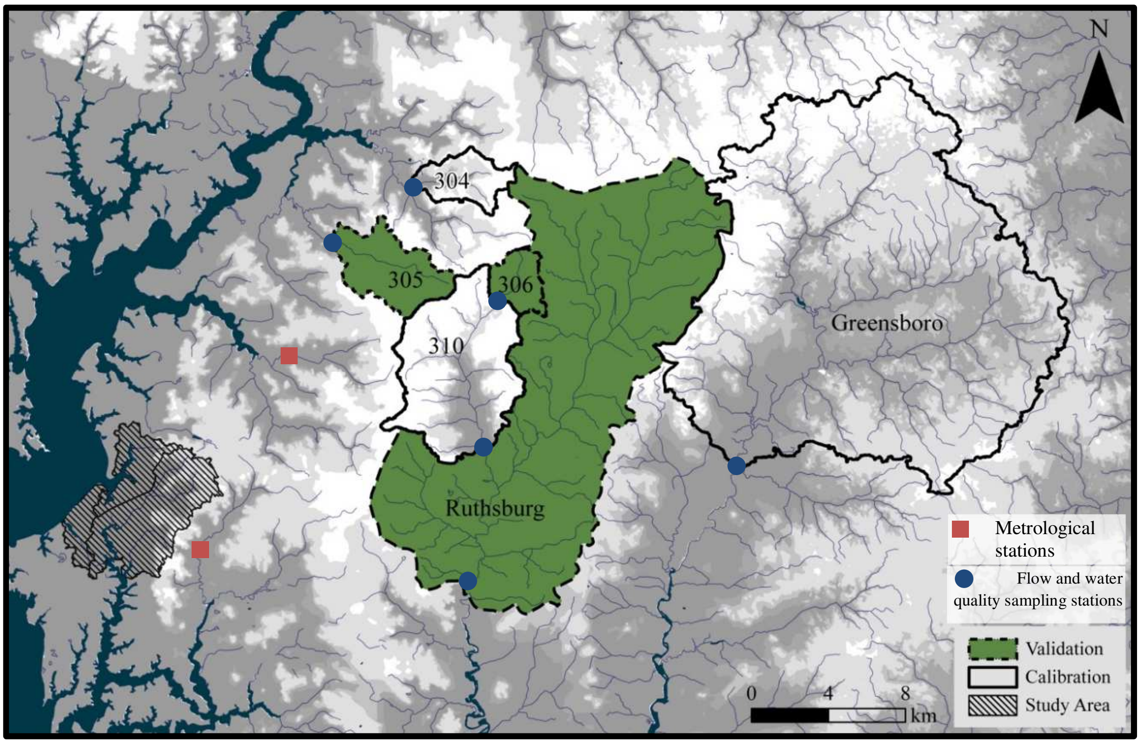

Figure 3. Location of the calibration and validation watersheds near the Queenstown study area (hatched shading).

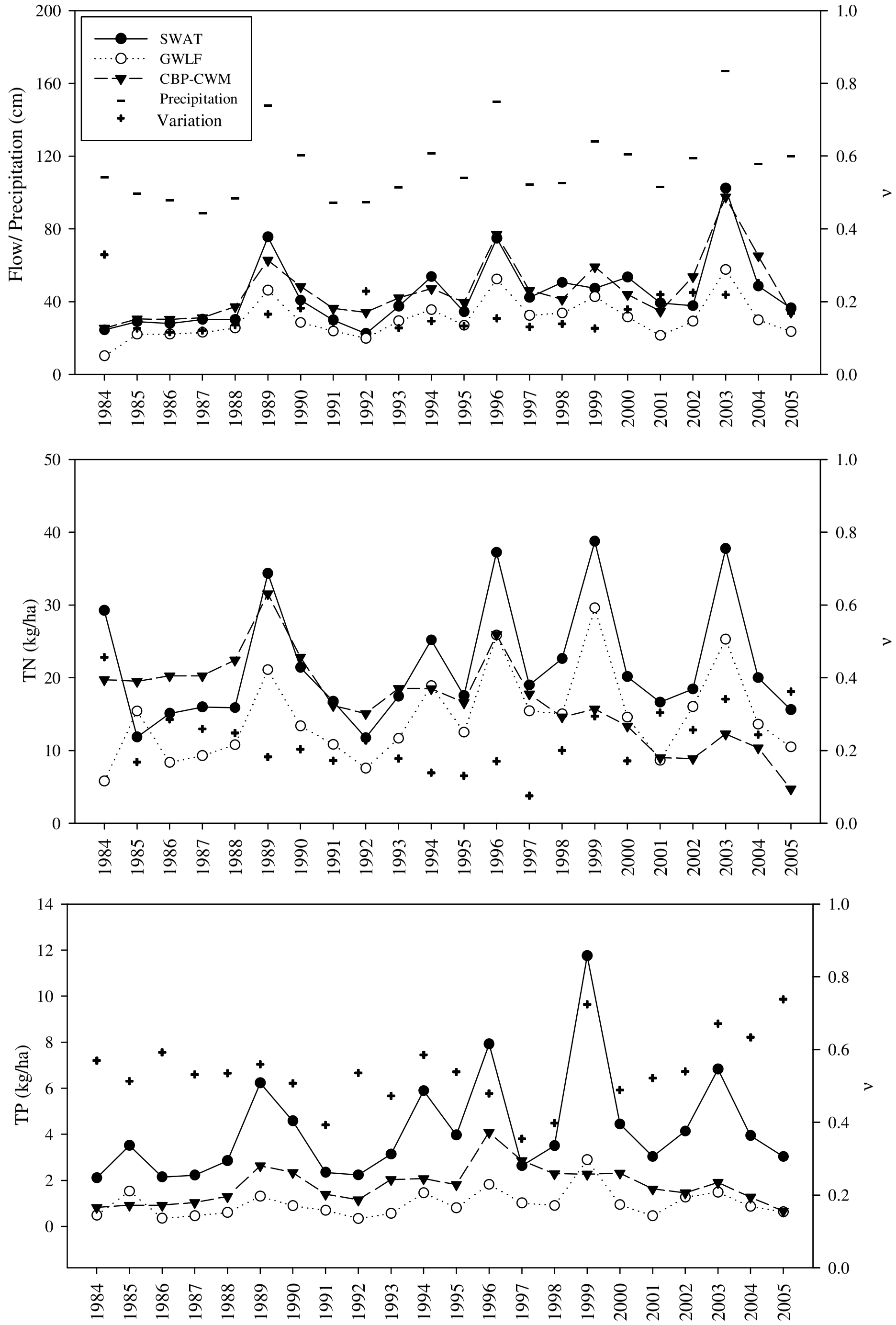

Figure 4. Annual flow and nutrient loadings predicted by each model under current land use for Queenstown watershed. Variation index (ν) is shown on right axes.

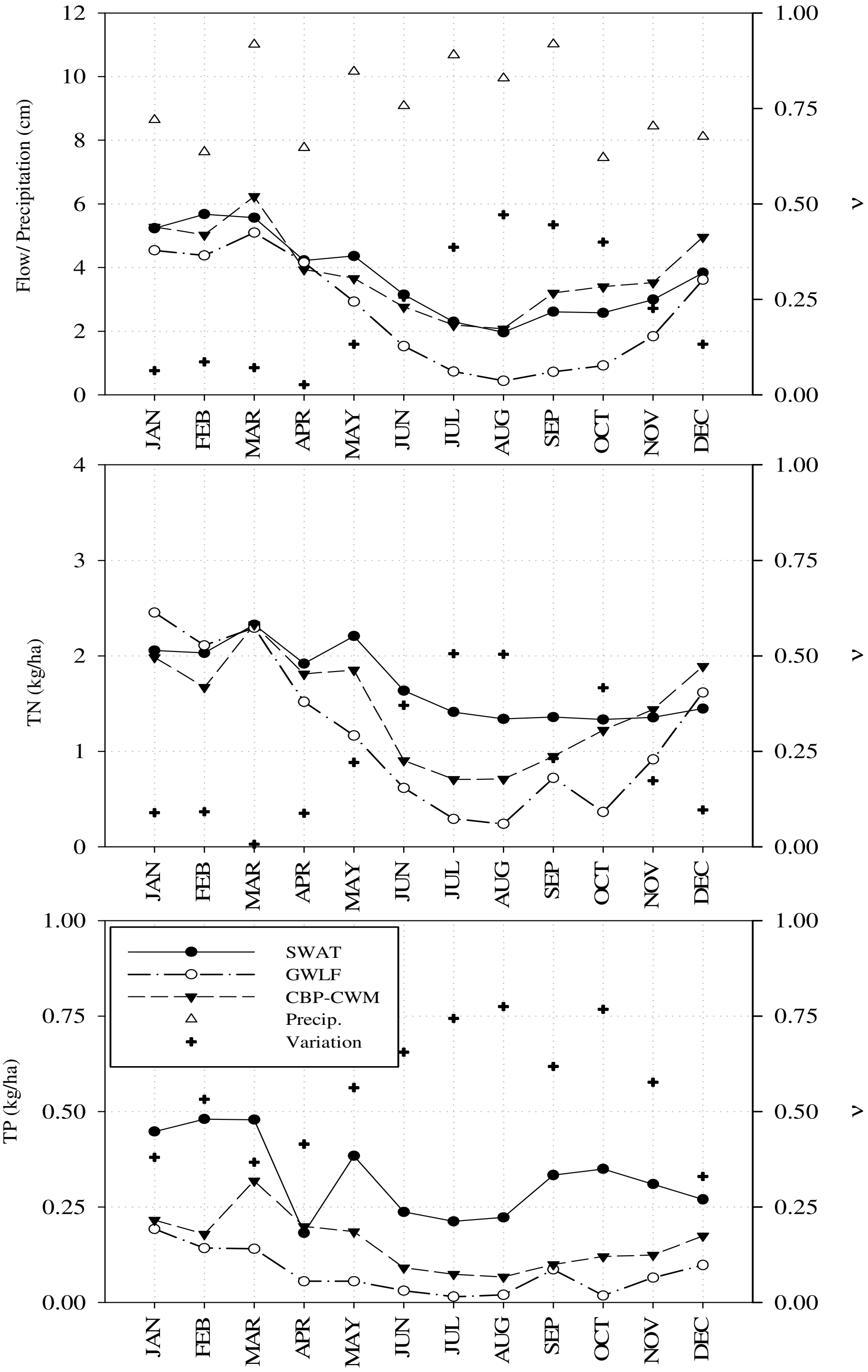

Figure 5. Average monthly flow and nutrient loadings predicted by each model under current land use for Queenstown watershed (1984-2005). Variation index (ν) is shown on right axes.

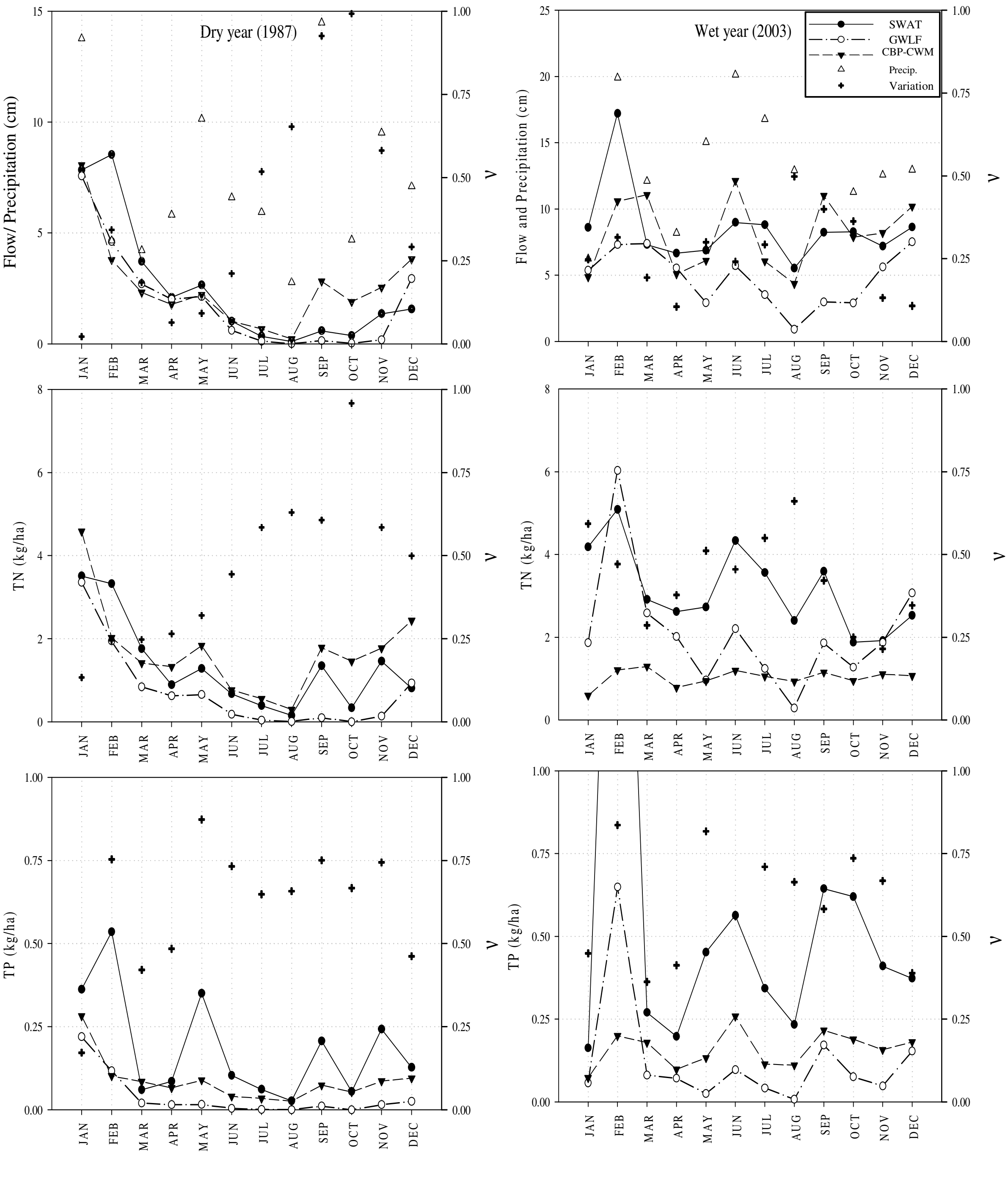

Figure 6. Flow and nutrient loading predictions by models for a dry year (1987, left) and a wet year (2003, right) under current land use. The results are combined for the three watersheds

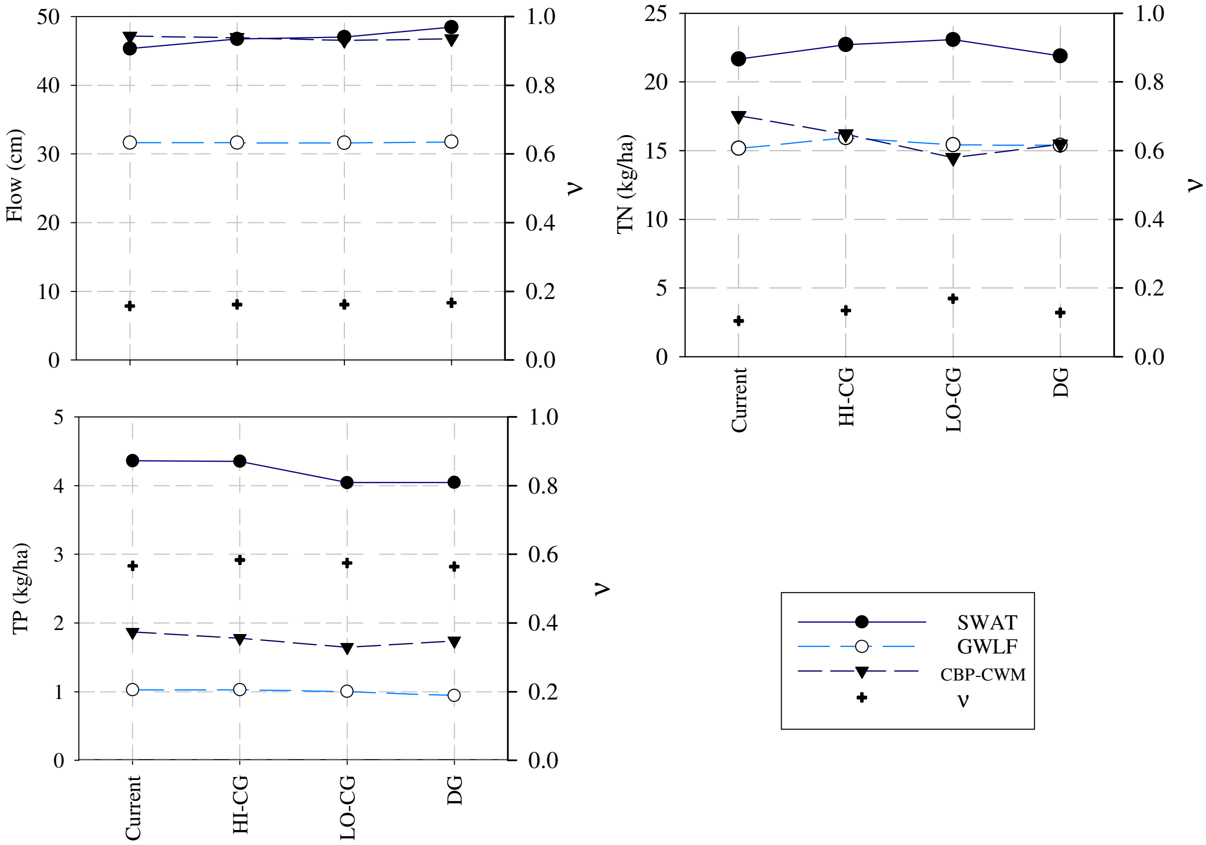

Figure 7. Average annual flow and nutrient load predictions by each model under different land use scenarios for the three watersheds combined

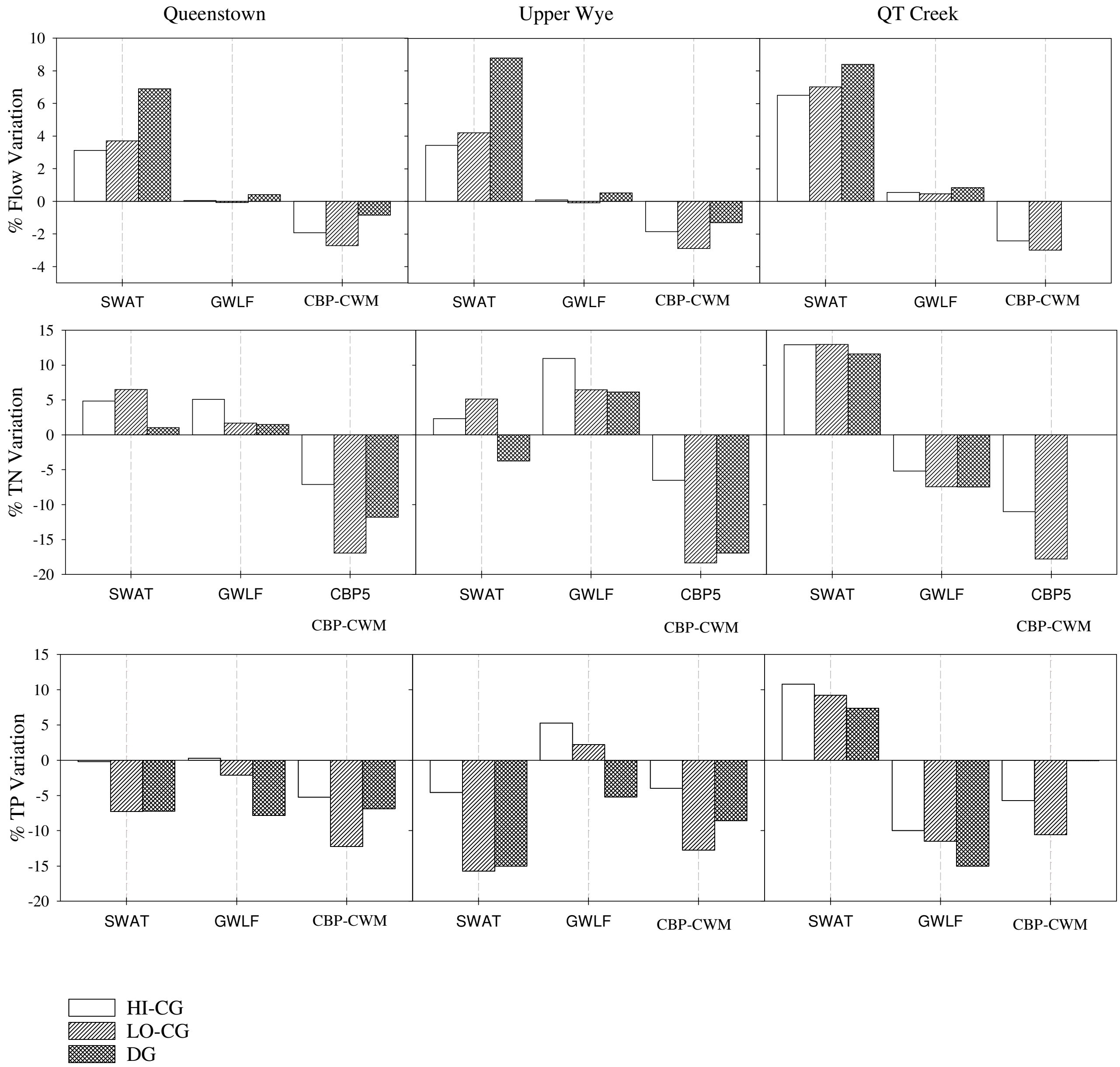

Figure 8. The percentage variation of mean annual flow, TN and TP fluxes at Queenstown, Upper Wye and QT Creek watersheds compared to current land use simulations, predicted by each model

Appendix

Figures A1 to A6 present comparison of observed and model predicted monthly streamflow, TP and TN fluxes at calibration (304, 310 and Greensboro) and validation (306, 306 and Ruthsburg) watersheds.

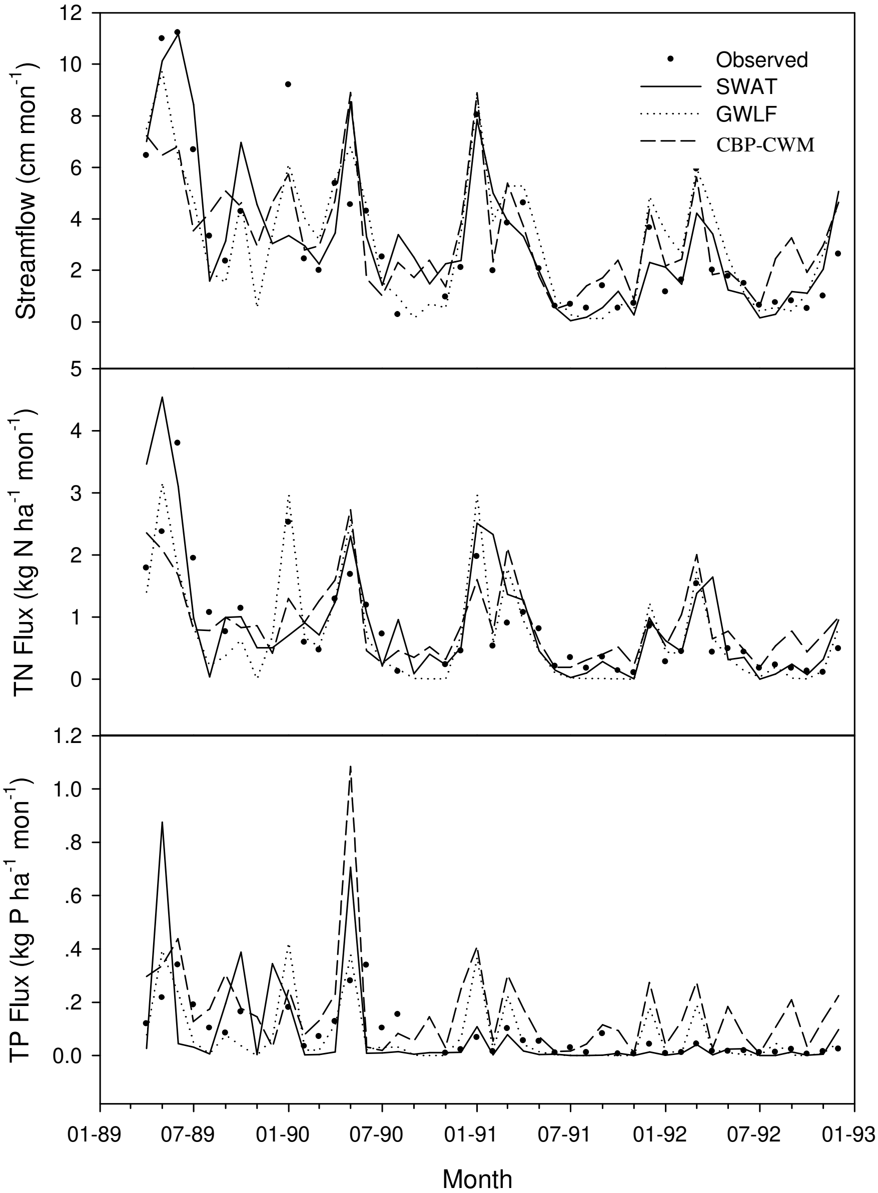

Figure A 1: Comparison of observed and model predicted monthly streamflow, TN and TP fluxes at watershed 304. Calibration period is April 1989 ~ Dec. 1992.

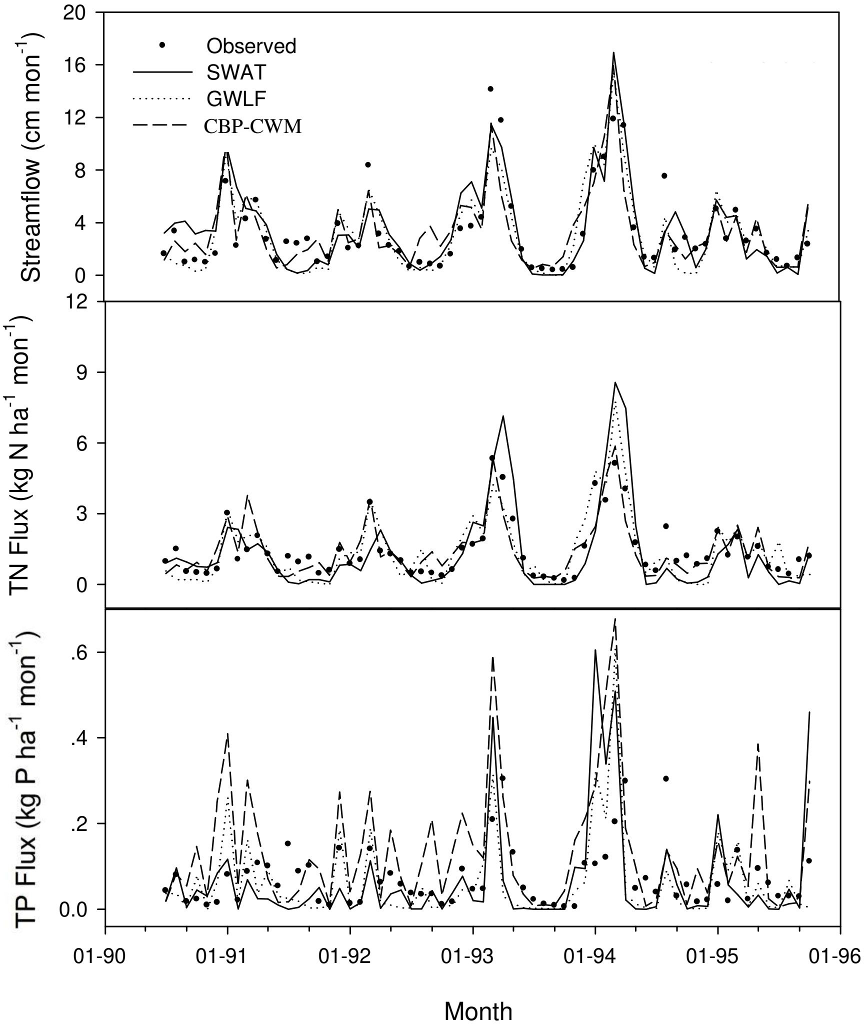

Figure A 2: Comparison of observed and model predicted monthly streamflow, TN and TP fluxes at watershed 310. Calibration period is July 1990 ~ Oct. 1995.

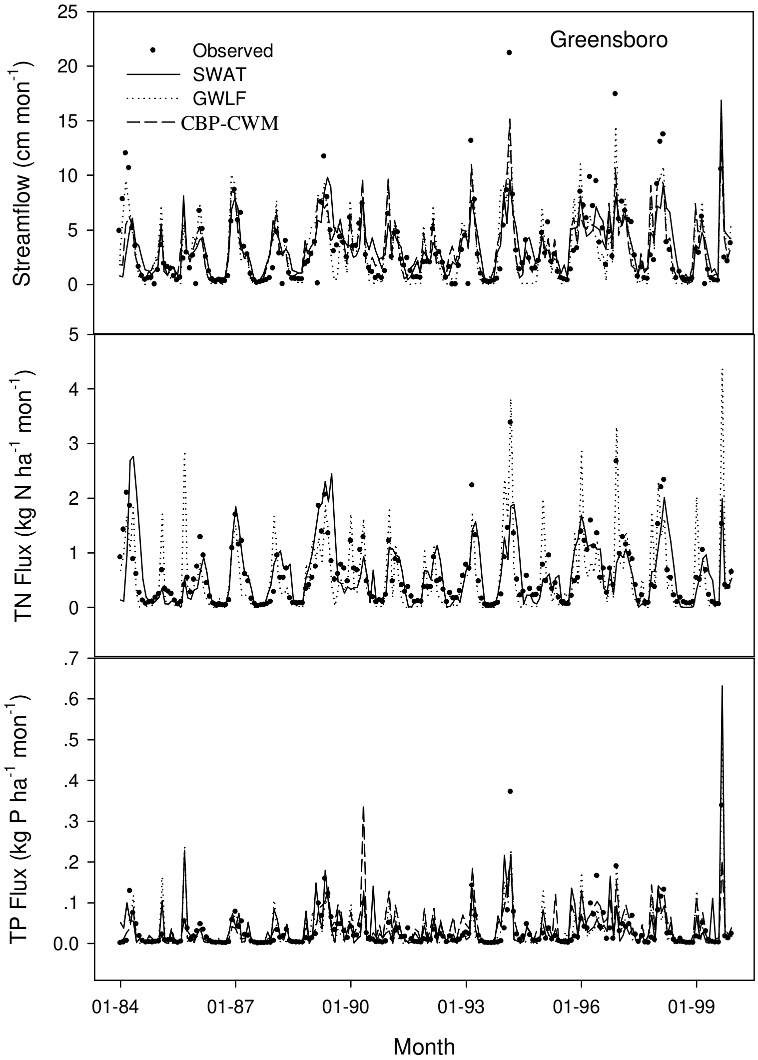

Figure A 3: Comparison of observed and model predicted monthly streamflow, TN and TP fluxes at watershed Greensboro. Calibration period is Jan. 1984 ~ Dec. 1999.

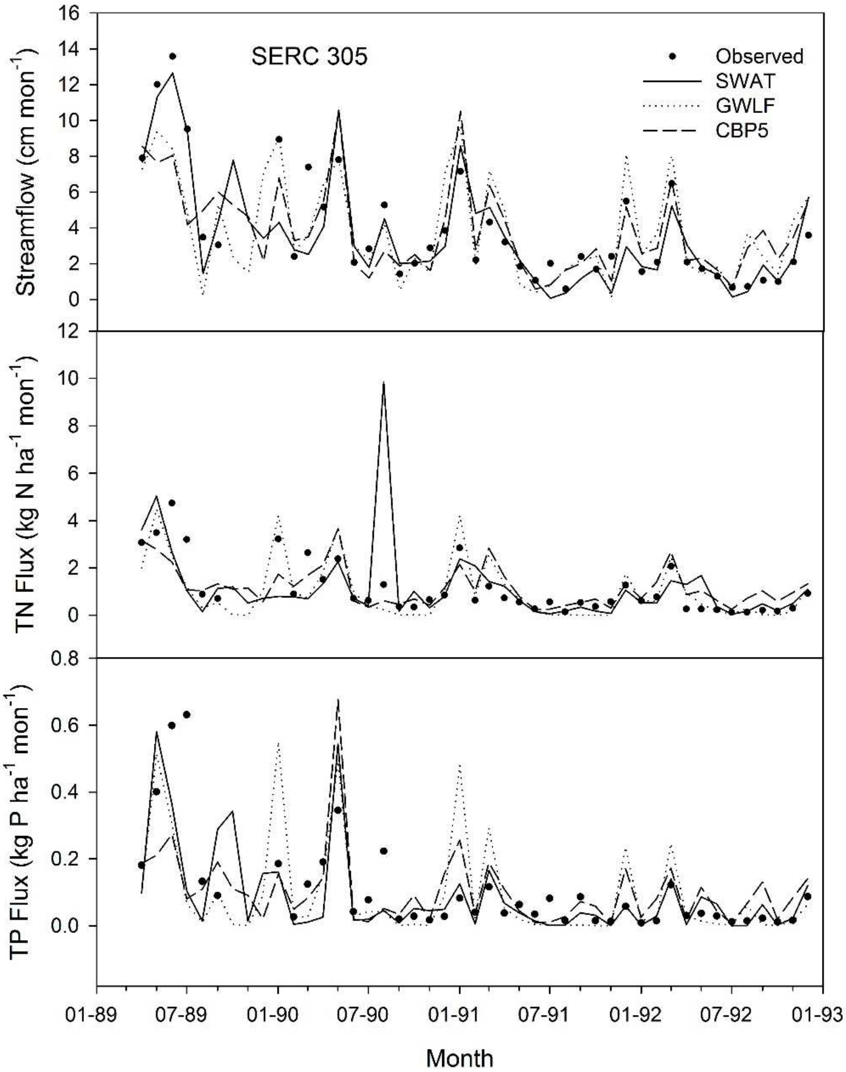

Figure A 4: Comparison of observed and model predicted monthly streamflow, TN and TP fluxes at watershed 305. Validation period is April 1989 ~ Dec. 1992.

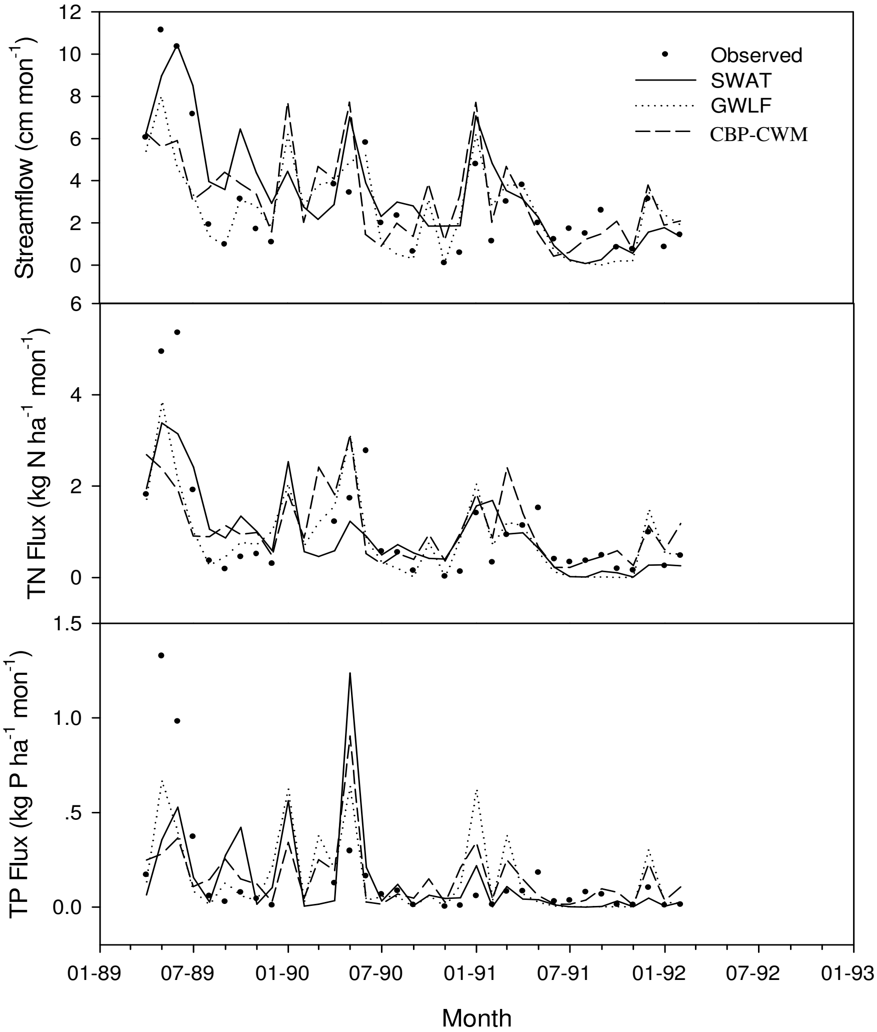

Figure A 5: Comparison of observed and model predicted monthly streamflow, TP and TN fluxes at watershed 306. Validation period is April 1989 ~ Feb. 1992.

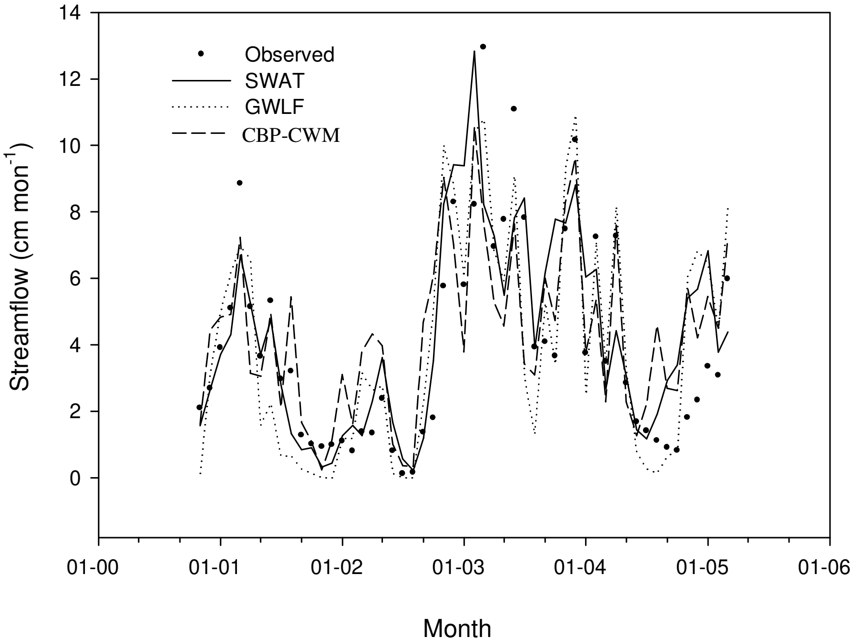

Figure A 6: Comparison of observed and model predicted monthly streamflow at watershed Ruthsburg. Calibration period is Nov 2000 ~ March. 2005.