Improving Model Prediction Reliability through Enhanced Representation of Wetland Soil Processes and Constrained Model Auto Calibration – A Paired Watershed Study

Amirreza Sharifi ^{a,*}, Megan W. Lang ^b, Gregory W. McCarty ^c, Ali M. Sadeghi ^c, Sangchul Lee ^b, Haw Yen ^d, Martin C. Rabenhorst ^e, Jaehak Jeong ^d and In-Young Yeo ^f

^a Department of Environmental sciences and Technology, University of Maryland, 10300 Baltimore Ave. BLDG. 007, RM. 104, BARC-WEST, Beltsville MD 20705

^b Department of Geographical Sciences, University of Maryland, 10300 Baltimore Ave. BLDG. 007, RM. 104, BARC-WEST, Beltsville MD 20705

^c Hydrology and Remote Sensing Laboratory, USDA-ARS, 10300 Baltimore Ave. BLDG. 007, RM. 104, BARC-WEST, Beltsville MD 20705

^d Blackland Research and Extension Center, Texas A&M University, 720 East Blackland Road, Temple, TX 76502, USA

^e Department of Environmental sciences and Technology, University of Maryland, 0208 H. J. Patterson Bldg., College Park, MD 20742

^f School of Engineering, The University of Newcastle, EA 127, University Drive, Callaghan, NSW 2308, Australia

* Corresponding author at: Amirreza Sharifi, Post-doctoral Research Associate, Department of Environmental sciences and Technology, University of Maryland, 10300 Baltimore Ave. BLDG. 007, RM. 104, BARC-WEST, Beltsville MD 20705, USA. Tel.: +1 (301) 504 6074.
E-mail address: sharifi@umd.edu

Abstract

Process based, distributed watershed models possess a large number of parameters that are not directly measured in field and need to be calibrated, in most cases through matching modeled in-stream fluxes with monitored data. Recently, concern has been raised regarding the reliability of this common calibration practice, because models that are deemed to be adequately calibrated based on commonly used metrics (e.g. Nash Sutcliffe efficiency) may not realistically represent intra-watershed responses or fluxes. Such shortcomings stem from the use of an evaluation criteria that only concerns the global in-stream responses of the model without investigating intra-watershed responses. In this study, we introduce a modification to the Soil and Water Assessment Tool (SWAT) model, and a new calibration technique that collectively reduce the chance of misrepresenting intra-watershed responses. The SWAT model was modified to better represent NO_3 cycling in soils with various degrees of water holding capacity. The new calibration tool has the capacity to calibrate paired watersheds simultaneously within a single framework. It was found that when both proposed methodologies were applied jointly to two paired watersheds on the Delmarva Peninsula, the performance of the models as judged based on conventional metrics suffered, however, the intra-watershed responses (e.g., mass of NO_3 lost to denitrification) in the two models automatically converged to realistic sums. This approach also demonstrates the capacity to spatially distinguish areas of high denitrification potential, an ability that has implications for improved management of prior converted wetlands under crop production and for identifying prominent areas for wetland restoration.

Keywords: Paired Watersheds, denitrification, SWAT-PMCT, Choptank River watershed, prior-converted cropland, hydric soils

1. Introduction

Process-based, distributed watershed models have been increasingly used in recent years for assessment of environmental policies and best management practices, both challenging tasks [Arnold et al., 2015]. Policy makers and stakeholders often make important decisions based on predicted model responses and perhaps more critically for jurisdictional regulation of nonpoint source pollution associated with agriculture. It is common knowledge that distributed, process-based models have a large number of parameters, partly due to the complex nature of the physical and biogeochemical processes that they characterize, and partly in an attempt to represent watershed spatial heterogeneity [Whittaker et al., 2010]. Most of these parameters are not directly measured *in situ* and need to be calibrated, in most cases through matching modeled in-stream fluxes (e.g. streamflow and nutrient loads) with monitored data [Beven, 2000; Yen et al., 2014a]. The calibration process is either performed manually, or using an autocalibration program that employs an optimization scheme for maximizing an objective function, which is based on statistics that reflect goodness of fit between model results and field observations. In this conventional calibration practice, it is assumed that the model reflects true system behavior when the “global” model responses at the outlet of the watershed (in-stream fluxes) adequately match the field observations [Yen et al., 2014a]. Adequacy of this match is subjective; however, there are publications, such as Moriasi et al. [2007], which offer ranges of acceptable goodness of fit statistics, measured using Nash-Sutcliffe efficiency and percent bias for flow and other predicted water quality constituents.

Recently, considerable concern has been raised regarding the reliability of conventional calibration practices [Arnold et al., 2015; White et al., 2014; Yen et al., 2015a]. This concern stems from the fact that so called adequately calibrated models (given the conventional standards

described above) may contain input data errors not readily identifiable by model users [White et al., 2014], or may not realistically represent intra-watershed responses or fluxes (e.g., crop yield, decomposition, denitrification, and NO₃ leaching) [Yen et al., 2014a]. Such shortcomings are a result of the large number of calibration parameters, the broad range that each parameter adopts, and the use of an evaluation criteria that is based solely on global in-stream responses of the model and not intra-watershed responses [Cassidy and Jordan, 2011; Yen et al., 2014a]. In other words, conventionally calibrated models might be producing results that perfectly match in-stream field observations, and at the same time over/under estimate intra-watershed responses by several orders of magnitude. Such flaws can be substantial, and lead to erroneous predictions when models are used to predict effects of conservation programs, pollution control strategies, best management practices or future climate scenarios. In response to the shortcomings of conventional calibration methods, some techniques and recommendations have been developed and suggested by the scientific community. White et al., [2014] developed a screening tool named “SWAT Check” that monitors SWAT model [Soil and Water Assessment Tool – Arnold et al., 2012] outputs and alerts users if simulated budgets (i.e., water, sediment and nutrient) are not realistic and outside of typical ranges. Yen et al. [2014a] incorporated extra measures in their autocalibration routine to check for some intra-watershed responses (i.e., denitrification and nitrogen lost to tile flow) and penalize simulations that produced unrealistic responses. The result was a calibrated model that not only matched available in-stream monitoring data, but produced realistic internal watershed behavior. Arnold et al. [2015] reviewed calibration strategies of 25 model application studies at different scales and provided recommendations for calibration/validation of watershed models. The recommendations include a four step process that embraces use of “hard data” – measured discharge and water quality at the outlet of a

watershed – and “soft data” – estimations of intra-watershed responses and other average estimations of physical evidence, such as groundwater depth, baseflow ratios, etc. [Yilmaz et al., 2008; Seibert and McDonnell, 2002].

In this study, we introduce a modification to the SWAT model, and a new calibration technique that collectively reduce the chance of misrepresenting intra-watershed responses, especially those related to wetlands. The proposed modification to the SWAT model concerns the resolution at which nutrient biogeochemical exchanges occur. In particular, this study focused on the simulation of denitrification – one of the major mechanisms for removal of NO_3 in terrestrial ecosystems. In the SWAT model, denitrification occurs at each soil layer when soil moisture of that layer is above a certain threshold value; and the amount of NO_3 lost to denitrification is calculated as a function of NO_3 concentration, soil organic carbon content and a global (basin wide) denitrification rate which is adjusted for local temperature effects. With this current SWAT protocol minimal distinction is made between poorly drained hydric soils that likely have higher potential for denitrification and better drained soil types. The modification that we propose in this study includes expanding the influence of the denitrification coefficient, by assigning various denitrification rates for various hydrologic response units as opposed to one rate for the whole basin, and allocating higher rates for soils with greater denitrification potential. Soil denitrification potential is extracted from a readily available source (i.e. the Soil Survey Geographic Database [SSURGO]).

Furthermore this study introduces a new calibration technique that has the capacity to calibrate paired watersheds simultaneously using a single framework. The term “paired watersheds” typically refers to a case where two (or more) independent watersheds are monitored for

116 discharge and water quality. Paired watersheds are usually within the same physiographic region,
117 are more or less about the same size, have many similar properties, but possess some distinctive
118 characteristics that make the comparison of their outflows compelling. The purpose of
119 monitoring paired watersheds is to understand how differences in soil properties, physiographic
120 attributes, or land management practices affect hydrology and nutrient cycling within the paired
121 watersheds. The common practice of modeling paired watersheds involves the construction of
122 individual models for each watershed, which are calibrated independently without paying
123 attention to the shared characteristics of the paired watersheds. In the new method proposed in
124 this study, the calibration of paired watersheds is not performed independently, but on one
125 platform that uses an optimization algorithm for autocalibration. In this new approach, the user
126 identifies calibration parameters that should be held constant between the two watershed models,
127 according to the shared properties of the two watersheds. The rest of the parameters are allowed
128 to adopt independent values. Accordingly, the autocalibration code generates one value for
129 shared parameters, and two values for independent parameters.

130 The objectives of this study were to : (1) introduce a modification to the SWAT model that
131 enhances the representation of soils at different hydric states without introducing new
132 parameters; (2) introduce and test a new calibration technique for paired watershed experiments,
133 that constrains model parameters which represent shared processes between watersheds; and (3)
134 combine both methods to see whether implementing them in combination results in reasonable
135 estimates of intra-watershed responses using automated calibration.

2. Methods

2.1. Study area

The methodology proposed in this study was applied to two adjacent watersheds located within the upper region of the Choptank River basin, on the Delmarva Peninsula (Figure 1). The study watersheds, namely Greensboro and Tuckahoe, are roughly the same size and have similar crops and agricultural management practices, but possess distinctive soil properties. The Greensboro watershed (290.1 km²) originates in Kent County, Delaware and extends southwest towards the township of Greensboro, Maryland (Caroline County), where a U.S. Geological Survey (USGS) gauging station (#01491000) has been monitoring river stage since 1948. The Tuckahoe watershed (220.7 km²) is located within Queen Anne's and Caroline counties in Maryland, and flows south towards Tuckahoe State Park. USGS has been continuously monitoring flow at the outlet of the Tuckahoe watershed since 2000 (USGS Station # 01491500).

A combination of two facts has motivated extensive research and monitoring in the Choptank River basin in recent years [e.g. McCarty et al., 2008, Denver et al., 2014, de Guzman et al., 2012, Whitall et al., 2010; McCarty et al., 2014]: (1) the Choptank River has been classified as impaired under the Clean Water Act [USEPA, 2010], due partly to excessive nutrient and sediment loads from agriculture; and (2) the Choptank River basin has been designated as a study area for multiple federal studies, including the USDA Conservation Effects Assessment Project (CEAP)-Watersheds Component and the USDA Long-term Agro-ecosystem Research (LTAR) Network [NRCS 2010, McCarty et al., 2014, McCarty et al., 2008]. Roughly 60% of the Choptank River watershed is covered by agriculture; it is estimated that agriculture is responsible for over 80 % of Choptank watershed's nitrogen and phosphorus loading discharge to

Chesapeake Bay [Ator and Denver 2015]. Baseflow constitutes a significant part of streamflow in the Choptank River basin, and acts as a primary mechanism for delivery of nitrogen to the Chesapeake Bay. It is estimated that 43 % to 70 % of NO_3 flux in the region's surface waters is transported via base flow [Bachman et al. 1998; Sprague et al. 2000; Ator and Denver, 2012]. Average annual precipitation and temperature for the study region is 1123 mm and 13.3 °C, respectively.

The Greensboro and Tuckahoe watersheds possess qualities and characteristics that make them particularly suitable for this study. The two watersheds are adjacent, and adopt similar management practices. The Greensboro watershed is slightly larger than Tuckahoe (290.1 vs 220.7 km^2), and has a smaller percentage of agricultural land cover (36.1 % vs. 54.0 %) (Table 1). However, the total area of cropland in the two watersheds is comparable (104.4 km^2 in Greensboro vs. 119.1 km^2 in Tuckahoe). Knowing that agriculture is the main source of nutrient discharge in the Choptank River basin [Ator and Denver 2015], one would expect that our two study watersheds discharge the comparable amount of NO_3 at their outlets. However, this assumption is incorrect. Figure 2 illustrates that there is a twofold difference in the NO_3 loading between the two watersheds. This is most likely due to differences in watershed soil properties, as illustrated by the distribution of hydric soils within the two watersheds (Figure 3 and Table 2). A detailed look at soils of the Greensboro and Tuckahoe watersheds (Table 2) reveals that Greensboro has a higher percentage of poorly drained soils compared to Tuckahoe (62.5 % vs. 43.0 %). Perhaps more importantly roughly 49 % of agricultural lands in Greensboro have poorly drained soils as compared to 32 % of agricultural lands in Tuckahoe. Poorly drained soils often exhibit anaerobic conditions due to saturation. Compared to well drained soils, which typically remain aerobic, poorly drained soils have higher potentials for denitrification, and thus lower

181 NO₃ concentration assuming the same nitrogen inputs [Denver et al., 2014; Hively et al., 2011;
182 McCarty et al., 2014; Ator et al., 2015]. On the Delmarva Peninsula, the saturated conditions that
183 result in denitrification are common in current and even historical wetlands, including inland
184 depressional, flats, and riparian wetlands [Denver et al., 2014; Duff et al., 2008; Pucket et al.,
185 2008; Jordan et al., 2007], coastal swamps and marshes [Speiran, 1996], and along groundwater
186 flow paths in organic rich sediments [Bohlke and Denver, 1995; Denver et al., 2010]. The last
187 category is particularly important in the Choptank basin, where many wetlands have been
188 ditched and drained for agriculture [Norton and Fisher 2000]. Even when “drained”, the drainage
189 is at best only partially effective in lowering the water tables in the soils, and is less effective
190 with distance away from the drainage structure. The potential for denitrification in poorly
191 drained, hydric soils is likely higher than other soils types due to a higher chance of
192 anoxic/anaerobic conditions, a result of slow infiltration capacity and/or shallow water table
193 depths. Koskelo [2008] demonstrated that baseflow NO₃ concentrations on the Delmarva
194 Peninsula decreased with increasing area of hydric soils within forests, likely due to greater
195 denitrification in these areas. Similar findings were reported by Norton and Fisher [2000] who
196 found that forests with hydric soils showed a strong negative correlation with stream total
197 nitrogen (TN) and NO₃ in the Choptank basin. In a study reported by Hively et al. (2011) and
198 McCarty et al. (2014), monthly baseflow stream samples from 15 agricultural subwatersheds of
199 the Choptank River in Maryland USA (2005 to 2007) were analyzed for mean nitrate
200 concentrations. It was found that nitrate concentrations were greater (p=0.0001) in the well-
201 drained upland hydrogeomorphic region than in poorly drained upland, reflecting increased
202 denitrification and reduced agricultural land use intensity in the poorly drained upland landscape
203 due to the prevalence of hydric soils. Furthermore, nitrate was inversely correlated to percent

hydric soils of each subwatershed ($p < 0.001$), and a moderate (but significant) inverse relationship was observed between nitrate and cropland on hydric soil ($p < 0.01$). Based on the collected evidence, the authors concluded that smaller nitrate concentrations from poorly drained subwatersheds were most likely due to greater nitrate reduction within the subwatershed and not simply due to less cropland and therefore less nitrate-N application.

Denver et al. (2014) collected hydrologic, geochemical and water quality data from a set of prior converted croplands (currently farmed historical wetlands) for roughly 2 years (2007-2008) and observed that when the zone of reducing conditions associated with the historic wetland extended through the entire thickness of the surficial aquifer, the historic wetland was effective at reducing the overall flux of nitrate that passed through the wetland sediment.

On the Delmarva peninsula, the drainage classes assigned to soils identified during routine soil survey mapping activities are based upon the presence, quantity, type and location (depth) of particular soil morphological (redoximorphic) features. Emphasis is placed on the presence of “low chroma” redox depletions (what used to be called gray drainage mottles). These redox depletions are indicative prolonged saturation and the development of reducing (anaerobic) conditions that facilitate the reduction, mobilization and segregation of iron oxide coatings. In this region, these wet hydrological conditions (shallow water tables) are mostly present during the Winter and early Spring (when ET is low) and then the water tables usually drop during the late Spring and Summer (when ET is high.) The redoximorphic features commonly take a long time to form (possibly decades or centuries), and once formed, they also persist for a long time (possibly centuries or longer in the absence of disturbance). Thus, these features integrate and reflect the long term (or typical) hydrological condition in the soil (the morphology persists

regardless of whether a given year is a wet year or a dry year). In *well drained soils*, if present at all, these low-chroma redox depletions are absent from the upper meter of the soil. As soils become progressively more poorly drained, these depletions occur closer to the soil surface and are more abundant. Thus, *moderately well drained soils* have redox depletions between 50 and 100 cm from the soil surface. *Somewhat poorly drained soils* show redox depletions shallower than 50 cm. When the abundance of the gray depletions exceeds 50%, the soil horizon is named using a “g” subscript (i.e. Eg, Btg) and is described as having a “depleted matrix.” Thus, *poorly drained soils* have a depleted matrix that extends very near the soil surface (typically immediately below the A horizon). In particularly wet landscapes, discharging water (or tidal conditions) may cause the water table to persist at or near the soil surface for much longer periods, and the associated anaerobic conditions may inhibit and slow the decomposition of soil organic matter leading to the formation of thick, dark, A horizons (or in extreme cases, thick O horizons). Therefore, *very poorly drained soils* not only have a depleted matrix that extends very near the soil surface (like poorly drained soils), but also have a thick (20-25cm or more), dark (Munsell value/chroma 3/3 or darker – often 2/2 or darker) A or O horizon.

2.2. Soil and Water Assessment tool (SWAT)

The Soil and Water Assessment Tool (SWAT) is a semi-distributed, process-based, long-term hydrologic and water quality model [Neitsch et al., 2011] developed by the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS). SWAT was developed for assessing long-term impacts of management practices and non-point source pollutions in complex watersheds, and is a widely recognized tool for evaluating effectiveness of BMPs (best management practices)[Arnold et al., 2012]. The SWAT model has been widely applied across multiple water resource disciplines with over 2,200 peer reviewed publications [SWAT

Literature Database, 2015]. Major model components in SWAT include hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides, and agricultural management [Borah and Bera, 2003]. In the SWAT model, areas with unique land cover, slope, and soil attributes are called Hydrologic Response Units (HRU). HRUs are elementary simulation units and represent the scale at which many hydrologic and nutrient cycling processes are simulated [Yen et al., 2015b]. Three organic pools (residue, stable, and active nitrogen) and two inorganic pools (ammonia and NO_3) constitute different components of the Nitrogen cycle in the model, where biogeochemical processes such as mineralization, nitrification, denitrification, and volatilization govern the balance between different pools. Volumetric water content and average concentration of NO_3 in the soil govern NO_3 concentrations in runoff, lateral flow, and percolation [Neitsch et al., 2011]. A full description of model components and history of the SWAT model could be found in Arnold et al. [2012] and Neitsch et al. [2011].

2.2.1. Estimation of denitrification in SWAT

Denitrification is one of the primary mechanisms for removal of NO_3 in terrestrial ecosystems [Howarth et al., 1996]. In saturated, organic rich sediments, as oxygen is depleted during organic matter decomposition by microorganisms, anoxic/anaerobic conditions develop [Ator and Denver, 2015]. This lack of oxygen often results in the activity of other heterotrophic microorganisms, such as denitrifying bacteria that reduce NO_3 or NO_2^- molecules to gaseous end products [NO , N_2O or N_2]. In the SWAT model, denitrification only occurs within a soil layer when soil moisture is above a certain threshold [Neitsch et al., 2011]. Given the right moisture conditions, the amount of NO_3 lost to denitrification is calculated as a function of NO_3 concentration, soil organic carbon content and a global (basin wide) denitrification rate which is adjusted for local temperature effects, such that [Neitsch et al., 2011]:

$$N_{denit,ly} = NO_{3,ly} \times (1 - e^{-\beta \cdot \gamma_{ly} \cdot orgC_{ly}}) \quad (1)$$

where ly denotes soil layer, N_{denit} is the mass of nitrogen lost to denitrification [ML^{-2}], NO_3 is the mass of NO_3 per unit area in layer ly [ML^{-2}], β is the global rate exponential coefficient for denitrification [dimensionless], γ is the temperature adjustment factor for nutrient cycling in layer ly [dimensionless], and $orgC$ is the amount of organic carbon in the layer (%). There is minimal distinction made between different soil types when estimating mass losses for denitrification. The most influential calibration parameter adjusting denitrification in Eq. (1), β , is a global parameter, applied to all soils basin wide. This approach is problematic, as it lacks the capacity to account for higher denitrification losses in soils that exhibit more favorable conditions for anaerobic microbial activities, such as hydric soils.

2.2.2. Modification of SWAT source code for enhanced representation of denitrification

The SWAT model's source code was modified to address the problem stated above. In the updated SWAT code, various β values were assigned for different HRUs, the size and distribution of which are controlled by the user. In other words, in the modified code, β is no longer a global parameter. Rather it is disaggregated so that each HRU carries its own denitrification coefficient. Furthermore, higher β values are assigned to soils with higher denitrification potential. SSURGO, a nationally available soils database was used to identify soils with higher denitrification potential. In this study drainage classification was used to assign denitrification potential within the model, but other indicators could be used in future studies (e.g., hydrologic soil group, water table depth, flooding and ponding frequency, and hydric classification). SSURGO divides soils to six classes according to their drainage characteristics

(i.e., very poorly drained, poorly drained, somewhat poorly drained, moderately well drained, well drained, and excessively well drained) and three classes according to hydric soil status (i.e., hydric, partially hydric, and not hydric). Drainage class was selected as a surrogate for denitrification potential since it offered a more detailed division between soil groups. Consequently, β which can range from 0 to 3 [Neitsch et al., 2011], was divided into 6 ranges, and each range was assigned to a drainage class (Table 2). Well drained soils were assigned lower denitrification ranges whereas poorly drained soils were allocated higher ranges for β .

2.3. SWAT model parameterization

Multiple layers of data were used to parameterize the Greensboro and Tuckahoe SWAT models using the ArcSWAT GIS interface. Daily precipitation and temperature data, jointly shared between the two models, were downloaded from the National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (NCDC) for the Greensboro (US1MDCL0009 coordinates: 38.99°N, 75.78°W – precipitation only) and Easton-Newman field airport (72404303756, coordinates: 38.80°N, 76.07°W – temperature only) weather stations. A 10 m digital elevation model (DEM) was acquired from the USGS National Elevation Dataset [Gesch, 2009], and the soil map was downloaded from the USDA Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO). The detailed land use map employed in this study was produced using multiple layers of available land use data such USGS National Land Cover Database (NLCD), the USDA National Agriculture Statistics Service (NASS) National Cropland Data Layer (NCDL), and digital orthophoto quadrangle (DOQ) imagery from National Aerial Photography Program (NAPP) [Sexton et al., 2010]. NASS NCDL divides the region's cultivated area into 8 subcategories, based on their crop types and planting rotations (Table 3). As presented in Table 3, corn, soybean, and double crop winter

wheat/soybean are the crops/crop rotations found in the study area. Detailed agronomic management information for this study area is presented in

Table 4. This information was acquired through field investigations, literature review and interviewing experts. Daily stream flow data were acquired from USGS gauge stations on the Choptank River near Greensboro (USGS#01491000) and Tuckahoe Creek (USGS#01491500). Daily NO_3 data were acquired from USDA-ARS. Stream water NO_3 concentrations were measured by use of a full spectrum (200 to 700 nm) spectrometer probe (S-CAN spectrolyser probe, Vienna Austria) with 15 mm path length and fitted with a flow cell. Stream water was automatically sampled on a 30 minute interval by use of a peristaltic pump connected to the flow cell.

2.4. SWAT Autocalibration tool

In this article, we introduce a novel autocalibration tool named “SWAT-PMCT” (SWAT Paired Model Calibration Tool) that has the capacity to calibrate paired watersheds simultaneously within a single framework. As mentioned earlier, the common protocol for modeling paired watershed experiments involves constructing individual models for each watershed, and calibrating them independently. Such practice ignores the shared characteristics of the paired watersheds, which require identical representation (parameterization) in the models. SWAT-PMCT hires a single optimization algorithm to calibrate paired (two or more) watershed models on a single platform. Given the shared characteristics of paired watersheds, the user identifies a number of parameters that require identical representation in the paired models. In other words, the user identifies the constrained calibration parameters (i.e. the parameters that should have a similar value between the two models). Subsequently, SWAT-PMCT (1) provides one value for

constrained parameters, and two values for free parameters (one per model); (2) modifies model parameters within input files of both models and run models in parallel (simultaneously); and (3) evaluates a single objective function that reflects the goodness of fit between observed in-stream data and model results at the outlets of both watersheds. In this case, the objective function was derived from flow and NO₃ discharge values at the outlet of the Greensboro and Tuckahoe watersheds. A flow chart illustrating the parallel calibration approach in SWAT-PMCT can be seen in Figure 4. Needless to say that SWAT-PMCT could also be used to calibrate paired watersheds independently, by assigning zero constrained parameters.

The Dynamically Dimensioned Search method (DDS) [Tolson and Shoemaker, 2007] was used as the optimization algorithm in SWAT-PMCT. Briefly, the DDS is a popular global optimization approach used to conduct parameter estimation in the field of water resources engineering for low- and high-dimensional problems [Yen et al., 2014a and 2014b]. In a recent study, Yen et al., [2014a] compared the DDS method to a handful of other commonly utilized optimization techniques and concluded that in high-dimensional problems, DDS provides superior computational efficiency in finding better objective function values and error statistics. In the DDS method, users define the maximum number of function evaluations, and identify calibration parameters and allowable ranges for each parameter. DDS starts with a global search at the beginning, and becomes a more localized as the numbers of iterations reach the maximum allowable number of function evaluations, defined by the user [Tolson and Shoemaker, 2007]. In the case of this study, the maximum number of function evaluations was set to 3,000 simulations for all model runs.

2.5. Calibration strategies

Four different calibration strategies have been employed in this study (Table 5). In Case 1, the Greensboro and Tuckahoe models were calibrated individually using the DDS autocalibration tool. This is the traditionally used autocalibration approach, where the model is tuned to match observed in-stream data, which in this case is flow and NO₃ discharge values at the outlet of the watershed. For Case 2, a similar approach to Case 1 was employed, with the exception of how denitrification is simulated in the model. The SWAT model used in Case 2 was modified to reflect distributed denitrification at the HRU level, as explained earlier. In Case 3, the two models were calibrated in parallel, with the objective function derived from flow and NO₃ discharge at the outlets of Greensboro and Tuckahoe watersheds. For Case 4, a similar approach to Case 3 was employed, except that the distributed denitrification scheme was also applied. Flow diagrams illustrating the four calibration approaches are shown in Figure 4.

Calibration parameters and respected allowable ranges for each parameter are shared between all 4 calibration strategies (Table 6). Key calibration parameters and their allowable ranges were identified by consulting multiple regional studies that were performed on the Choptank River Basin (Chu et al., 2004; Zhang et al., 2008; Meng et al., 2010, Yen et al., 2015b and Yeo et al., 2015). The number of selected calibration parameters in this study (38 parameters) might appear excessive at first. However, one should note that yielding acceptable calibration results for nitrate loading at daily time scale is a relatively difficult task that requires a larger set of calibration parameters (compared with monthly calibration efforts). In this study, particular attention was paid to preserve the spatial distribution of the model. As shown by Table 6, most of the HRU and subbasin level parameters were only changed relative to their initial estimates, an

undertaking that guarantees conservation of the spatial heterogeneity of the model. Initial parameter estimates (for all model parameters) are drawn automatically using meteorological information, land use maps and soil layers, combined with internal model databases through a GIS interface, namely ArcSWAT (Zhang et al., 2013; Winchell et al., 2010)

For cases 3 and 4, more than half of the calibration parameters (20 parameters) are shared between the Greensboro and Tuckahoe models. These parameters were identified based on expert opinion. Through years of investigating Choptank River Watershed, we know that the substantial differences between the Greensboro and Tuckahoe watersheds boils down to their soil types. Given the fact stated above, most of the constrained parameters consist of the ones which control processes mainly unrelated to soil biogeochemistry. To be precise, the constrained parameters control physical processes related to snow hydrology (SMFMX, SMTMP, SNOCOV MX, TIMP), overland flow generation, routing and sediment transport (CN2, OV_N, CH_K1, SMXCO, ADJ_PKR), Ground water hydrology and nutrient transport (SHALLST, REVAPMN, NPERCO, HLIFE_NGW, LAT_ORGN), plant water uptake (EPCO), and some microbial constants (RSDCO, BC2_BSN, BC1_BSN, SDNCO, BC3_BSN). More than half of the calibration parameters are constrained between the two watersheds; this large number was selected in the interest of the idea that the more constrained the calibration is, more realistic results can be achieved.

There was only one year worth of daily NO_3 observations available for both watersheds, extending from April 1st, 2014 till April 6th, 2015. It was decided that such short period was only enough to calibrate models, thus, leaving no data for validation. Accordingly, a two year period

was designated for flow calibration, extending from April 1st, 2013 till April 6th, 2015, preceded by a 15 month warm up period.

2.6. Objective function

The objective function for the DDS algorithm was constructed using Nash-Sutcliffe efficiency (NSE) [Nash and Sutcliffe, 1970]. NSE can range from $-\infty$ to one, where NSE=1 represents a seamless match between observed and simulated data. NSE has become one of the most frequently used measures for estimating performance of various models (ASCE 1993), and is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i^{Obs} - y_i^{Sim})^2}{\sum_{i=1}^N (y_i^{Obs} - y_i^{Mean})^2} \quad (2)$$

where y_i^{Obs} is the observed data at time step i ; y_i^{Sim} is the simulated output at time step i ; y_i^{Mean} is the mean of observed data over all time steps; and N is the number of time steps. The objective function is characterized as the summation of NSE for two output variables, streamflow and NO_3 , such that:

$$OF = \sum_{v=1}^V (NSE_v) \quad (3)$$

where OF is the value of the objective function; NSE_v is the NSE for output variable v ; and V is the number of total output variables. For scenarios 1 and 2, $V=2$, as models are calibrated individually, only based on flow and NO_3 profiles at the outlet of each individual watershed. On the other hand, $V=4$ for scenarios 3 and 4, as there are 2 flow and 2 NO_3 profiles at the outlet of Greensboro and Tuckahoe watersheds.

2.7. Applications of behavior definition

The concept of behavior definition, proposed by Moriasi et al. [2007] for SWAT model applications, offers statistical thresholds characterized by conceptual indices (i.e., Very Good, Good, Satisfactory, and Unsatisfactory) as a useful supplementary tool in supporting the decision making processes. The threshold offered by Moriasi et al. [2007] has implications for simulations that are performed at the monthly scale. Generally, daily simulation goodness of fit statistics exhibit much higher performance when reassessed at monthly time scale. Therefore, thresholds of Moriasi et al. [2007] required some modifications before application within this study. The modified measures implemented in this study are presented in Table 7 [refer to Moriasi et al., [2007] for the original Table). The PBIAS section of Table 7 is identical to that of Moriasi et al. [2007], whereas the NSE section is different in two details: 1- the modified table has distinguished NSE thresholds for NO_3 and flow (the original method does not) 2- The NSE thresholds are toned down (more relaxed) in modified table – since performance measures will improve when reassessed at monthly scale. For instance, a “Very good” performance rating according to Moriasi et al. [2007] requires $0.75 < \text{NSE} < 1.00$, whereas in the modified ratings, it is altered to $0.65 < \text{NSE} < 1.00$.

3. Results and discussion

3.1. Model performance assessment

Figure 5 and Figure 6 show the time series of flow and NO_3 simulations in Greensboro and Tuckahoe watersheds for different case scenarios and field measured data. Table 8 presents model performance statistics for all calibration strategies. All case scenarios performed well for flow calibrations in both watersheds (Figure 5 and Table 8). Greensboro models consistently had

better flow calibration results compared to Tuckahoe models, with higher NSE ($0.65 < \text{NSE} < 0.66$ compared to $0.56 < \text{NSE} < 0.59$) and significantly lower $|\text{PBIAS}|$ ($0 < |\text{PBIAS}| < 5\%$ compared to $15 < |\text{PBIAS}| < 20\%$). Figure 5 shows that all models underestimate flow in dry periods (August to November) and fail to match observed flows in both watersheds, however, the problem is more pronounced for the case 3 Greensboro model. The reason behind this might be due to the fact that an objective function that relies on NSE will force the model to correct peak flow predictions, as Nash-Sutcliffe is not very sensitive to model overprediction or underprediction, specifically during low flow periods (Krause et al., 2005). For Tuckahoe, another source of uncertainty might be because of a 19 acre recreational lake, called Tuckahoe Lake that lies at the immediate upstream of USGS hydrologic monitoring gauge. This Lake is managed by Maryland Department of Natural Resources (MDNR), where lake levels, and downstream flows are adjusted seasonally. Based on our field observations, MDNR is releasing more flow during low flow seasons to maintain a healthy environmental flow, hence introducing errors in our model predictions during low flow season (i.e. Aug – Nov 2014).

To explore the effects of various calibration routines during high flows, Figure 7 displays model flow responses with varying probabilities of non-exceedance for different case scenarios and magnifies the area of high flow occurrences ($P > 95\%$). As shown in Figure 7 (bottom panels), for Greensboro all models tend to underestimate high flows, whereas in Tuckahoe models better matched observations for flows at high frequency of non-exceedance ($P > 95\%$). Tuckahoe models consistently overestimated a wide range of median flows (Figure 7; $0.25 < P < 0.95$), a fact that is reflected in Table 8 where all Tuckahoe models have a negative PBIAS for flow, with absolute values ranging from 15-20%.

It appears that applying calibration constraints (i.e., parallel calibration in cases 3 and 4) does not affect performance statistics for flow simulations. This suggests that the practice of parallel calibration has the benefit of minimizing unsubstantiated differences in intra-watershed processes without imposing negative effects on flow calibration.

Greensboro models performed considerably better in capturing NO_3 dynamics compared to Tuckahoe models (Table 8; $0.35 < \text{NSE} < 0.6$ compared to $0.31 < \text{NSE} < 0.34$); however, the difference is less pronounced for case 4 where their performance is comparable. The Greensboro model performed very well in case 3 ($\text{NSE} = 0.6$, $|\text{PBIAS}| = 14\%$); given that NO_3 calibration was performed at daily scale, this record indicates the robust abilities of SWAT-PMCT. When the new denitrification scheme was applied along with the parallel calibration approach (case 4), Greensboro model performance suffered ($\text{NSE} = 0.35$). Tuckahoe model performance statistics varied minimally between different case scenarios ($0.31 \leq \text{NSE} \leq 0.34$, $18\% \leq |\text{PBIAS}| \leq 25\%$). NSE statistics for Tuckahoe NO_3 predictions improved when reassessed at monthly scale – a common scale for nutrient flux performance estimations – with $0.53 \leq \text{NSE} \leq 0.60$, thus putting Tuckahoe models within to the satisfactory state, defined by Moriasi et al. [2007]. All models fail to capture NO_3 dynamics over winter of 2015, when observed NO_3 fluxes were highly dynamic (Figure 6). Tuckahoe models consistently overestimated NO_3 fluxes until mid-July of 2014, and then underestimated NO_3 export until the end of the calibration period. Tuckahoe models considerably underestimate high NO_3 fluxes (Figure 8). Therefore all Tuckahoe models have a low NSE score since NSE is sensitive to high extreme values and tends to penalize simulations that fail to capture peaks [Moriasi et al., 2007]. Figure 8 and Table 8 both point to the fact that all models in both watersheds underestimate NO_3 fluxes during the calibration period.

Comparing NO₃ performance statistics for various calibration cases, it was observed that model performance for Greensboro was improved noticeably when the parallel scheme was applied (case 1 vs. case 3), and only suffered minimally in the case of Tuckahoe; hence advocating the use of parallel calibration practice in paired watersheds. Another observation was that the case 4 model of Greensboro (denitrification scheme + parallel calibration) has inferior performance compared to other case scenarios, indicating that the denitrification scheme does not necessarily improve model performance.

3.2. Assessment of intra-watershed responses

The main objective of this study was to assess the effects of different calibration approaches on intra-watershed responses, and to determine whether applying the introduced methodologies in combination leads to more reasonable model estimation of intra-watershed responses. It is a general rule of thumb that out of many different responses that mechanistic models predict, reliable responses are the ones associated with field observations that were used to calibrate/validate the model. In this study, daily NO₃ data was used to calibrate the models, thus two important internal watershed responses associated with NO₃ cycling (i.e., NO₃ yield from groundwater and NO₃ lost to denitrification) were investigated for different case scenarios (Table 9). Note that intra-watershed responses are difficult to measure, and field measured balances are rarely available [White et al., 2014]. Therefore, the validity of information provided in Table 9 was examined by comparing the budgets of Greensboro and Tuckahoe watersheds with respect to each other, and against the amount of fertilizer applied to agricultural fields. To appropriately compare the two watersheds, balances presented in Table 9 were estimated for agricultural HRUs only.

507 An interesting observation from Table 9 is the large discrepancy between amounts of
508 denitrification reported by various Greensboro and Tuckahoe scenarios. Case 3 models report the
509 least amount of denitrification among all models (2.4 and 2.1 kg N ha⁻¹ yr⁻¹ for Greensboro and
510 Tuckahoe, respectively). This amount is not realistic given the amount of fertilizer applied to
511 agricultural fields in both watersheds (~100 kg-N ha⁻¹ yr⁻¹). The root of this problem is likely
512 parameter estimations that the autocalibration program converged on (Table 6). As mentioned
513 earlier, apart from the available NO₃ in the soil, denitrification in the SWAT model is a function
514 of two calibration parameters: CDN (denitrification exponential rate coefficient – β in (1)) and
515 SDNCO (denitrification threshold water content). The final value for both of these parameters in
516 the case 3 models are too extreme (Table 6). SDNCO has converged on a very high value (i.e.,
517 0.993), and CDN on a small value (i.e., 0.010 for Greensboro and 0.020 for Tuckahoe), resulting
518 in minimal amounts of denitrification in agricultural HRUs. However, case 3 models have the
519 strongest combined performance for NO₃ calibration according to conventional metrics. But
520 given that the representation of denitrification in both models is not realistic, one might advise
521 against using case 3 model for future scenario analysis. This demonstrates that conventionally
522 used model performance metrics do not always indicate model performance quality. Case 1
523 models estimate denitrification values that look valid if considered independently (48.5 and 19.2
524 kg N ha⁻¹ yr⁻¹ for Greensboro and Tuckahoe, respectively), but such variation between
525 Greensboro and Tuckahoe denitrification losses is unrealistic. As was mentioned earlier the two
526 watersheds have differences in their soil properties and distribution, but that difference cannot
527 explain a twofold variance in observed denitrification losses. Case 2 and Case 4 models estimate
528 similar denitrification losses for the two watersheds (i.e., 42.8 and 40.5 kg N ha⁻¹ yr⁻¹ in case 2;
529 36.7 and 33.8 kg N ha⁻¹ yr⁻¹ in case 4 for Greensboro and Tuckahoe, respectively). Not only were

all the suggested estimations by case 2 and case 4 models within the accepted range ($\leq 42\%$ of applied fertilizer), but unlike case 1, the difference between estimated denitrification losses between the two watersheds seem reasonable as well. Greensboro exhibited higher denitrification losses compared to Tuckahoe, likely due to its higher percentage of poorly drained soils in agricultural areas. Case 2 models had better performance statistics than case 4 models, enough to make models of case 2 more appealing than case 4 models. However, calibration strategy was more constraining for case 4 (parallel calibration + distributed denitrification), insuring more consistent results between the Greensboro and Tuckahoe models. The more constrained calibration logically filters out simulations that perform well in one watershed, but not in the other – as the objective function of autocalibration is comprised of performance statistics in both watersheds. This practice also tends to weed out simulations that demonstrate good performances for the wrong reasons (i.e., unrealistic intra- watershed responses). Case 2 models are calibrated independently, lacking all the mechanisms for weeding out corrupted simulations. Given the facts stated above, it is more logical to put one's trust in case 4 models, despite their weaker performances. Case 4 models reflect the distributed denitrification scheme, and share many common calibration parameters, a measure that maintains the concord between Greensboro and Tuckahoe models, thus adding confidence to their predictions.

In regards to NO_3 yield from groundwater balance (Table 9) there was strong agreement between all case scenario models that the Tuckahoe watershed receives higher loads of NO_3 from groundwater than Greensboro. A highly likely explanation for this two to threefold difference is lower denitrification rates in the Tuckahoe watershed leading to a buildup of NO_3 in groundwater. The difference between Greensboro and Tuckahoe's denitrification balance as

reported by case 4 models was $\sim 3 \text{ kg N ha}^{-1} \text{ yr}^{-1}$, enough to explain the buildup of NO_3 in Tuckahoe's groundwater system since the invention of industrial fertilizers.

3.3. Autocalibration of denitrification coefficients in case 2 and case 4

Table 10 presents autocalibrated rates of denitrification coefficient – β in Eq. (1) – for soils of various drainage classes in cases 2 and 4. Note that min and max bounds for these rates were subjectively assigned, with 0.5 units distance between β of various categories of soils. Assigning strict boundaries between different categories can be problematic in cases. For instance, in both case 2 and 4, the β for “Poorly drained” soils converged to a value close to 2.000, which is the lower bound for β in that category of soils (Table 2). This means that the autocalibration routine would probably have adopted a lower value for β if allowed, and the overall model performance may have been improved. One possible solution is to assign flexible boundaries for β , i.e. allowing β in different categories to adopt any value within the larger allowable range ($0 < \beta < 3$), as long as the requirement of preeminence (i.e. $\beta_{\text{poorly drained soils}} > \beta_{\text{well drained soils}}$) is satisfied.

3.4. Convergence patterns of autocalibration objective function

As mentioned earlier, the autocalibration of all scenarios was conducted for 3,000 model evaluations per case scenario. To see how different calibration techniques affected autocalibration results, the value of objective function versus model iterations for each scenario at the two watersheds was illustrated in Figure 9. The objective function used for autocalibration was the summation of NSE for flow and NO_3 . Note that for cases 3 and 4, the objective function value is composed of Greensboro + Tuckahoe NSE results. For all cases in both watersheds, major improvement was found within the first 1,000 iterations; after that, the objective function

only improved incrementally (Figure 9). Figure 9 also illustrated that case 3 and 4 models take more iterations to stabilize compared to case 1 and 2. For instance, case 1 and 2 models for Tuckahoe stabilized around 800 iterations, whereas case 3 and 4 models did not improve noticeably until after 2,200 iterations. This is not surprising since cases 3 and 4 are more complex parallel scenarios that require more iterations to be stabilized. Accordingly, 3,000 iterations was found to be sufficient for achieving the best possible outcome for all calibration scenarios.

3.5. Assessment of behavioral simulations

All 3,000 iterations for each case scenario were assessed against behavior definitions described in Table 7, and the percentage of behavior solutions are shown in Figure 10 for both watersheds. Looking at the Greensboro models (left panel in Figure 10), case 3 was the only model that produced “Very good” results in both watersheds. Case 4, which had the most calibration restrictions (parallel + denitrification scheme), had no “Very good” or “Good” results. Based on Greensboro results, the parallel calibration practice improved model performance (case 1 vs. case 3). However, the comparison between cases 3 and 4 revealed that application of the modified denitrification scheme (on the top of parallel calibration scheme) did not help improve model performance statistics.

None of the Tuckahoe models produced “Very good” or “Satisfactory” results (right panel in Figure 10). This was expected, since the Tuckahoe models did not have as sound performance statistics as Greensboro models (Table 8). Case 3 (parallel calibration) produced the weakest results with only 26% “Satisfactory” runs. The most “Satisfactory” results were produced by case 1 followed by case 2 and case 4 (61%, 57% and 47%, respectively). Case 3 and case 4

models were both calibrated in parallel; an act that caused the internal model responses in both watersheds to be in balance with each other, while case 1 and 2 models were calibrated independently from each other, allowing models more freedom to exhibit good performances, but not necessarily for the right reasons (with reasonable intra-watershed responses).

3.6. Distribution of denitrification at field scale

One of the main advantages of the modified denitrification scheme is the enhanced spatial representation of denitrification in the landscape. Figure 11 shows an aerial image and a soil map of a typical agricultural field in the Greensboro watershed. The poorly drained soils in the field have reduced crop biomass – as seen on the aerial photograph. The bottom panels of Figure 11 show the spatial representation of denitrification loss rates estimated by case 1 and case 2 models. According to data in Table 9, denitrification loss in Greensboro models of case 1 and case scenarios were within the ballpark of each other (48.47 and 42.76 kg N ha⁻¹ yr⁻¹, respectively). When the denitrification loss was mapped over all HRUs, it was evident that the case 2 model clearly distinguishes areas of high denitrification potential (hot spots), whereas the case 1 model does not.

4. Summary and Conclusion

This paper introduces a modification to the SWAT model that improves the spatial representation of soils at various degrees of wetness, and a new calibration technique that has the capacity to autocalibrate paired watersheds simultaneously within a single framework. The primary goal of these modifications was to enhance model representation and understanding of biogeochemical processes which influence water quality, specifically those that are largely controlled by wetland related processes. This was partly achieved by spatially disaggregating

denitrification through the assignment of different denitrification rates to wetland and non-wetland soils within the basin. Assigning denitrification rates based on propensity for this process using a readily available data source, such as SSURGO, allows the application of this method throughout the US, and anywhere else in the world with sufficient soil, hydroperiod, or wetland maps. New wetland mapping techniques will further improve the ability to designate areas that are likely to support different levels of denitrification [Huang et al. 2014; Lang et al. 2012; Lang et al. 2009; Lang et al. 2008]. Our goal was also supported by the introduction of a new parallel calibration technique, which calibrates multiple models within a single platform to ensure that parameters that are shared between watersheds in the real world are also shared within the modeling framework. Although it is possible that doing so may reduce model performance as judged using traditional metrics, a more realistic assessment of model uncertainty is produced when real-world conditions are allowed to guide model processes. Furthermore, this protocol supports the use of modeled results to improve the understanding of environmental process – and not simply to mimic stream outflows. This is an important step towards the enhanced use of models to guide environmental decision-making.

When both of the aforementioned methodologies were applied to the Greensboro and Tuckahoe watershed models model performance suffered. However, intra-watershed responses in the Greensboro and Tuckahoe models automatically converged on realistic sums. The dual watershed autocalibration routine helped to eliminate simulations that performed well in one watershed but not in the other, thus disallowing simulations that demonstrated good model performance for the wrong reasons (i.e., unrealistic intra-watershed responses). Applying the parallel calibration scheme did not affect performance statistics for flow simulations in either watershed.

Models with a modified denitrification scheme were able to clearly distinguish areas of high denitrification potential (hot spots or critical areas). This spatially explicit information can be used to not only improve the understanding of environmental processes and interactions with agronomic practices, but can also be used to guide the implementation of conservation practices. For example, the functioning of these critical zones could be enhanced through the placement of controlled drainage structures to encourage soil saturation, and thus denitrification, as cropping schedules allow. Findings from the Mid-Atlantic Regional Conservation Effects Assessment Project (CEAP-Wetlands) and the Choptank CEAP-Watersheds Benchmark watershed studies have demonstrated that currently farmed historical wetlands (i.e., prior converted croplands), are likely to have a considerable effect on water quality in the Mid-Atlantic Coastal Plain [Denver et al., 2014; McCarty et al., 2014; McCarty et al., 2008; Kluber et al., 2014; Ducey et al., 2015]. The ability to enhance scientific understanding of these areas at the watershed scale is vital to improving water quality in the Chesapeake Bay and elsewhere.

Our results support the *in situ* studies cited above, by confirming the importance of farmed historical wetlands, and the conceptual model proposed by McCarty et al. [2014]. In brief, this conceptual model emphasizes the importance of poorly drained soils on croplands determining the extent of nitrate processing within agricultural watersheds. This relationship follows from the inherently close association of the nitrate source from crop production and the soils comprising the cropland and both the intrinsic (soil) and extrinsic (landscape) properties of poorly drained soils that tend to promote nitrate processing in agricultural landscapes. Models consistently showed that Tuckahoe watershed received higher loads of NO_3 from groundwater than Greensboro. This phenomenon was explained by Tuckahoe's lack of ability to denitrify

excessive NO₃ (compared to Greensboro), hence causing a buildup of NO₃ in the groundwater system over the years.

This paper describes the development and initial testing of enhanced modeling protocols, designed to improve the understanding of intra-watershed responses. Additional work is needed to further refine and test these adaptations. For instance, future efforts should include the validation of model results. Validation helps to identify and eliminate overfitting during calibration [Whittaker et al., 2010]. The Greensboro and Tuckahoe watersheds would be well suited for future validation efforts because daily NO₃ values are currently being collected at the outlet of each watershed. Only 12 months of daily NO₃ data were available when this study was performed, but these data will be collected into the future as part of a CEAP-Watersheds Benchmark Watershed study and the Lower Chesapeake Long-term Agroecosystem Research site. Another area that calls for further examination is the effect of number, and specifications of constraint parameters on the performance of parallel autocalibration. In this study, the constrained parameters (listed in Table 6) were identified using expert opinion, and some iterative trials. A systematic approach for identifying such parameters would further improve parallel autocalibration protocols. Lastly, in this study SSURGO drainage class was used as a surrogate for denitrification potential. Other attributes in SSURGO (e.g., hydrologic soil group, water table depth, flooding and ponding frequency, and hydric classification) or other maps entirely may also be well suited for this purpose.

Acknowledgments

This project was funded by the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Conservation Effects Assessment Project (CEAP)

685 Wetland and Watershed Components. The authors would like to thank Mari-Vaughn Johnson,
686 USDA NRCS, for her support and constructive review of the manuscript.

687 DISCLAIMER

688 Mention of specific products is for identification and does not imply endorsement by the US
689 Department of Agriculture to the exclusion of other suitable products or suppliers.

References

- Arnold, J., Moriasi, D., Gassman, P., Abbaspour, K., White, M., Srinivasan, R., Santhi, C., Harmel, R., Van Griensven, A., Van Liew, M., 2012. SWAT: Model use, calibration, and validation. *T Asabe* 55, 1491-1508.
- Arnold, J.G., Youssef, M.A., Yen, H., White, M.J., Sheshukov, A., Sadeghi, A.M., Moriasi, D.N., Steiner, J.L., Amatya, D.M., Haney, E.B., Jeong, J., Arabi, M., Gowda, P.H., 2015. Hydrological Processes and Model Representation. *American Society of Agricultural and Biological Engineering* (Accepted - In Press).
- ASCE Task Committee, A., 1993. The ASCE task committee on definition of criteria for evaluation of watershed models of the watershed management committee Irrigation and Drainage Division, Criteria for evaluation of watershed models. *J. Irri. Drain. Engng*, ASCE 119, 429-442.
- Ator, S.W., Denver, J.M., 2012. Estimating Contributions of Nitrate and Herbicides From Groundwater to Headwater Streams, Northern Atlantsdic Coastal Plain, United States. *J Am Water Resour As* 48, 1075-1090.
- Authro, 2015. Understanding nutrients in the Chesapeake Bay watershed and implications for management and restoration: the Eastern Shore. 2330-5703.
- Bachman, L.J., Lindsey, B.D., Brakebill, J.W., Powars, D.S., 1998. Ground-water discharge and base-flow nitrate loads of nontidal streams, and their relation to a hydrogeomorphic classification of the Chesapeake Bay Watershed, middle Atlantic coast. US Department of the Interior, US Geological Survey.
- Beven, K.J., 2011. Rainfall-runoff modelling: the primer. John Wiley & Sons.
- Böhlke, J., Denver, J., 1995. Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland. *Water Resources Research* 31, 2319-2339.
- Borah, D., Bera, M., 2003. Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Transactions of the ASAE* 46, 1553-1566.
- Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: comparison with near-continuous phosphorus time-series data. *Journal of Hydrology* 405, 182-193.
- Chu, T.W., Shirmohammadi, A., Montas, H., Sadeghi, A., 2004. Evaluation of the SWAT model's sediment and nutrient components in the piedmont physiographic region of Maryland. *Transactions of the Asae* 47, 1523-1538.

Denver, J., Ator, S., Lang, M., Fisher, T., Gustafson, A., Fox, R., Clune, J., McCarty, G., 2014. Nitrate fate and transport through current and former depressional wetlands in an agricultural landscape, Choptank Watershed, Maryland, United States. *Journal of Soil and Water Conservation* 69, 1-16.

Denver, J.M., Tesoriero, A.J., Barbaro, J.R., 2010. Trends and transformation of nutrients and pesticides in a Coastal Plain aquifer system, United States. *Journal of environmental quality* 39, 154-167.

Authro, 2010. The quality of our Nation's waters-Nutrients in the Nation's streams and groundwater, 1992-2004. 2330-5703.

Ducey, T., Miller, J., Lang, M., Szogi, A., Hunt, P., Fenstermacher, D., Rabenhorst, M., McCarty, G., 2015. Soil Physicochemical Conditions, Denitrification Rates, and Abundance in North Carolina Coastal Plain Restored Wetlands. *Journal of Environmental Quality* 44, 1011-1022.

Duff, J.H., Tesoriero, A.J., Richardson, W.B., Strauss, E.A., Munn, M.D., 2008. Whole-stream response to nitrate loading in three streams draining agricultural landscapes. *Journal of Environmental Quality* 37, 1133-1144.

Gesch, D., Evans, G., Mauck, J., Hutchinson, J., Carswell Jr, W., 2009. The National Map-Elevation: US Geological Survey Fact Sheet 2009-3053, 4 p. Last accessible online at <http://ned.usgs.gov>.

Green, W.H., Ampt, G., 1911. Studies on soil physics, 1. The flow of air and water through soils. *J. Agric. Sci* 4, 1-24.

Hively, W.D., Hapeman, C.J., McConnell, L.L., Fisher, T.R., Rice, C.P., McCarty, G.W., Sadeghi, A.M., Whittall, D.R., Downey, P.M., de Guzmán, G.T.N., 2011. Relating nutrient and herbicide fate with landscape features and characteristics of 15 subwatersheds in the Choptank River watershed. *Science of the total environment* 409, 3866-3878.

Howarth, R.W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., Downing, J., Elmgren, R., Caraco, N., Jordan, T., 1996. Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences, Nitrogen cycling in the North Atlantic Ocean and its watersheds. *Springer*, pp. 75-139.

Huang, C.Q., Peng, Y., Lang, M.G., Yeo, I.Y., McCarty, G., 2014. Wetland inundation mapping and change monitoring using Landsat and airborne LiDAR data. *Remote Sens Environ* 141, 231-242.

Hunt, P., Miller, J., Ducey, T., Lang, M., Szogi, A., McCarty, G., 2014. Denitrification in soils of hydrologically restored wetlands relative to natural and converted wetlands in the Mid-Atlantic coastal plain of the USA. *Ecological Engineering* 71, 438-447.

Jordan, T.E., Andrews, M.P., Szuch, R.P., Whigham, D.F., Weller, D.E., Jacobs, A.D., 2007. Comparing functional assessments of wetlands to measurements of soil characteristics and nitrogen processing. *Wetlands* 27, 479-497.

Kellogg, D., Gold, A., Groffman, P., Addy, K., Stolt, M., Blazejewski, G., 2005. In situ ground water denitrification in stratified, permeable soils underlying riparian wetlands. *Journal of Environmental Quality* 34, 524-533.

Kluber, L.A., Miller, J.O., Ducey, T.F., Hunt, P.G., Lang, M., S. Ro, K., 2014. Multistate assessment of wetland restoration on CO₂ and N₂O emissions and soil bacterial communities. *Applied Soil Ecology* 76, 87-94.

Koskelo, A.I., 2008. Hydrologic and biogeochemical storm response in Choptank Basin headwaters. University of Maryland, College Park, Md. :.

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

Lang, M., McCarty, G., Oesterling, R., Yeo, I.Y., 2013. Topographic Metrics for Improved Mapping of Forested Wetlands. *Wetlands* 33, 141-155.

Lang, M.W., Kasischke, E.S., Prince, S.D., Pittman, K.W., 2008. Assessment of C-band synthetic aperture radar data for mapping and monitoring Coastal Plain forested wetlands in the Mid-Atlantic Region, USA. *Remote Sens Environ* 112, 4120-4130.

Lang, M.W., McCarty, G.W., 2009. Lidar Intensity for Improved Detection of Inundation Below the Forest Canopy. *Wetlands* 29, 1166-1178.

Lee, S., Yeo, I.-Y., Sadeghi, A.M., McCarty, G.W., W.D., H., Lang, M.W., 2015. Impacts of Watershed Characteristics and Crop Rotations on Winter Cover Crop Nitrate Uptake Capacity within Agricultural Watersheds in the Chesapeake Bay Region. *PLOS ONE* (under review).

McCarty, G., McConnell, L., Hapeman, C., Sadeghi, A., Graff, C., Hively, W., Lang, M., Fisher, T., Jordan, T., Rice, C., 2008. Water quality and conservation practice effects in the Choptank River watershed. *Journal of soil and water conservation* 63, 461-474.

McCarty, G.W., Hapeman, C.J., Rice, C.P., Hively, W.D., McConnell, L.L., Sadeghi, A.M., Lang, M.W., Whittall, D.R., Bialek, K., Downey, P., 2014. Metolachlor metabolite (MESA) reveals agricultural nitrate-N fate and transport in Choptank River watershed. *Science of the Total Environment* 473, 473-482.

Meng, H., Sexton, A.M., Maddox, M.C., Sood, A., Brown, C.W., Ferraro, R.R., Murtugudde, R., 2010. Modeling Rappahannock River Basin Using Swat - Pilot for Chesapeake Bay Watershed. *Appl Eng Agric* 26, 795-805.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. Asabe* 50, 885-900.

Nash, J., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology* 10, 282-290.

Authro, 2011. Soil and water assessment tool theoretical documentation version 2009.

Norton, M., Fisher, T., 2000. The effects of forest on stream water quality in two coastal plain watersheds of the Chesapeake Bay. *Ecological Engineering* 14, 337-362.

Puckett, L.J., Zamora, C., Essaid, H., Wilson, J.T., Johnson, H.M., Brayton, M.J., Vogel, J.R., 2008. Transport and fate of nitrate at the ground-water/surface-water interface. *Journal of environmental quality* 37, 1034-1050.

Seibert, J., McDonnell, J.J., 2003. The quest for an improved dialog between modeler and experimentalist. *Water Science and Applications* 6, 301-316.

Sexton, A.M., Sadeghi, A.M., Zhang, X., Srinivasan, R., Shirmohammadi, A., 2010. Using Nexrad and Rain Gauge Precipitation Data for Hydrologic Calibration of Swat in a Northeastern Watershed. *T Asabe* 53, 1501-1510.

Authro, 1996. Geohydrology and geochemistry near coastal ground-water-discharge areas of the Eastern Shore, Virginia.

Authro, 2000. Factors affecting nutrient trends in major rivers of the Chesapeake Bay watershed.

SWAT Literature Database, S., 2015. SWAT Literature Database for Peer-Reviewed Journal Articles.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* 43.

United States Department of Agriculture Natural Resources Conservation Service, N., 2010. Conservation Effects Assessment project. Washington, DC: US Department of Agriculture.

United States Department of Agriculture Soil Conservation Service, S., 1972. National Engineering Handbook, Section 4: Hydrology. Washington, DC.

United States Environmental Protection Agency, U., 2010. Mid-Atlantic water, 303(d) lists and 305(d) reports. United States Environmental Protection Agency - Washington DC.

White, M.J., Harmel, R.D., Arnold, J.G., Williams, J.R., 2014. SWAT check: A screening tool to assist users in the identification of potential model application problems. *Journal of environmental quality* 43, 208-214.

Whittaker, G., Confesor, R., Di Luzio, M., Arnold, J.G., 2010. Detection of Overparameterization and Overfitting in an Automatic Calibration of Swat. *T Asabe* 53, 1487-1499.

Winchell, M., Srinivasan, R., Di Luzio, M., Arnold, J., 2011. ArcSWAT interface for SWAT2009, user's guide < <http://swatmodel.tamu.edu/software/arcswat/>>. Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX; Grassland. Soil and Water Research Laboratory, USDA Agricultural Research Service, Temple, TX.

Yen, H., Bailey, R.T., Arabi, M., Ahmadi, M., White, M.J., Arnold, J.G., 2014a. The role of interior watershed processes in improving parameter estimation and performance of watershed models. *Journal of environmental quality* 43, 1601-1613.

Yen, H., Jeong, J., Tseng, W.H., Kim, M.K., Records, R.M., Arabi, M., 2015a. Computational Procedure for Evaluating Sampling Techniques on Watershed Model Calibration. *Journal of Hydrologic Engineering* 20.

Yen, H., Sharifi, A., Kalin, L., Mirhosseini, G., Arnold, J.G., 2015b. Assessment of model predictions and parameter transferability by alternative land use data on watershed modeling. *Journal of Hydrology* 527, 458-470.

Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014b. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. *Environmental Modelling & Software* 54, 211-221.

Yeo, I.Y., Lee, S., Sadeghi, A.M., Beeson, P.C., Hively, W.D., McCarty, G.W., Lang, M.W., 2014. Assessing winter cover crop nutrient uptake efficiency using a water quality simulation model. *Hydrol Earth Syst Sc* 18, 5239-5253.

Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research* 44.

Zhang, X., Srinivasan, R., Van Liew, M., 2008. Multi-Site Calibration of the Swat Model for Hydrologic Modeling. *T Asabe* 51, 2039-2049.

Zhang, X.S., Beeson, P., Link, R., Manowitz, D., Izaurralde, R.C., Sadeghi, A., Thomson, A.M., Sahajpal, R., Srinivasan, R., Arnold, J.G., 2013. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environmental Modelling & Software* 46, 208-218.

Tables

Table 1: Land use distribution over Greensboro and Tuckahoe watersheds

Land use	Greensboro	Tuckahoe
Agriculture	36.1 %	54.0 %
Forest	48.3 %	32.8 %
Pasture	9.3 %	8.4 %
Urban	5.6 %	4.2 %
Water body	0.7 %	0.6 %

Table 2: Amount of different soil drainage classes within the Greensboro and Tuckahoe watersheds (Middle two columns). Last two columns show upper and lower bounds for denitrification rate (β) assigned to soils in the study region according to their respected drainage class

Soil drainage class	% Greensboro	% Tuckahoe	Lower bound for β	Upper bound for β
Excessively well drained	3.0 (1.8)*	0.2 (0.0)	0	0.5
Well drained	17.7 (28.9)	36.2 (48.4)	0.5	1
Moderately well drained	16.8 (20.8)	20.6 (19.4)	1	1.5
Somewhat poorly drained	3.0 (3.8)	0.5 (0.3)	1.5	2
Poorly drained	34.2 (29.3)	31.5 (25.4)	2	2.5
Very poorly drained	25.3 (15.4)	11.0 (6.5)	2.5	3

* The values in the parenthesis denote the proportion of that soil type covering agricultural lands

Table 3: Agricultural subcategories represented in the land use map. Each subcategory has a unique crop rotation and management schedule.

Agricultural subcategory	2012	2013	2014	2015
AGA1	Dbl WW/Soy	Corn	Dbl WW/Soy	Dbl WW/Soy
AGA2	Corn	Dbl WW/Soy	Corn	Corn
AGAB	Soybean	Corn	Dbl WW/Soy	Soybean
AGB1	Soybean	Corn	Soybean	Soybean
AGB2	Corn	Soybean	Corn	Corn
AGC1	Corn	Corn	Corn	Corn
AGD1	Corn	Soybean	Soybean	Corn
AGD2	Soybean	Corn	Soybean	Soybean

Table 4: Management schedules for the Greensboro and Tuckahoe watersheds.

Crop	Planting	Fertilizer	Harvest
Corn	Apr. 30	157 kg N·ha ⁻¹ of poultry manure on Apr. 20 45 kg N·ha ⁻¹ of sidedress 30% UAN on Jun. 7	Oct. 3
Soybean	May 20		Oct. 15
Double crop winter wheat	Oct. 10	34 kg N·ha ⁻¹ of sidedress 30% UAN on Oct. 8 45 kg N·ha ⁻¹ of sidedress 30% UAN on Mar. 1 67 kg N·ha ⁻¹ of sidedress 30% UAN on Apr. 5	Jun. 27
Double crop soybean	Jun. 29		Nov. 1

Table 5: Calibration schemes employed in this study

	Description
Case 1	Greensboro and Tuckahoe models are calibrated individually
Case 2	Same as Case 1, except that the distributed denitrification scheme is applied
Case 3	Greensboro and Tuckahoe models are calibrated in parallel mode
Case 4	Same as Case 3, except that distributed denitrification scheme is applied

Table 6: Calibration parameters and their best fitted values for different calibration strategies.

	Parameter	Level	Min	Max	Unit	Description	Constrained? ¥	Case 1		Case 2		Case 3		Case 4	
								G	T	G	T	G	T	G	T
1	ADJ_PKR	Basin	0.5	2		Peak adjustment factor for sediment routing	Yes	0.517	0.554	0.576	0.525	0.507		0.503	
2	EPCO	Basin	0.5	1		Plant uptake compensation factor	Yes	0.794	0.978	0.719	0.826	0.538		0.712	
3	ESCO	Basin	0.1	1		Soil evaporation compensation factor		0.510	0.843	0.720	0.944	0.228	0.751	0.599	0.693
4	SURLAG	Basin	0.001	15	day	Surface runoff lag time		11.721	0.634	9.882	8.575	12.878	13.636	0.731	8.894
5	SMFMX	Basin	0	10	mm/Cd	Maximum melt rate for snow during year	Yes	7.397	1.846	9.127	5.277	7.386		3.396	
6	SMTMP	Basin	-5	5	C	Snow melt base temperature	Yes	-4.387	3.712	-3.821	4.696	-4.867		-4.919	
7	SNOCVMX	Basin	0	500	mm	Minimum snow water content	Yes	1.194	6.036	55.311	9.889	106.12		18.235	
8	TIMP	Basin	0.01	1		Snow pack temperature lag factor	Yes	0.404	0.171	0.035	0.132	0.530		0.124	
9	SDNCO	Basin	0	1		Denitrification threshold water content	Yes	0.579	0.417	0.254	0.356	0.993		0.976	
10	BC3_BSN	Basin	0.02	0.4	1/day	Rate hydrolosis of organic N to ammonia	Yes	0.205	0.070	0.039	0.046	0.025		0.265	
11	MSK_CO2	Basin	0	10		Coefficient impacting low flow		6.048	3.388	8.950	4.698	5.217	0.495	9.480	1.759
12	SMXCO	Basin	0	10		Adjustment for maximum curve number S factor	Yes	5.154	7.040	8.756	7.414	4.998		4.729	
13	CDN	Basin	0	3		Denitrification exponential rate coefficient		1.876	0.134	-	-	0.010	0.020	-	-
14	HLIFE_NGW	Basin	0	500	day	Half-life of nitrogen in groundwater	Yes	195.11	387.31	395.36	211.87	101.56		160.17	
15	NPERCO	Basin	0	1		Nitrogen percolation coefficient	Yes	0.677	0.959	0.842	0.939	0.868		0.856	
16	BC2_BSN	Basin	0.2	2	1/day	Rate constant for oxidation NO ₂ to NO ₃	Yes	0.446	0.728	1.812	1.840	0.874		0.939	
17	BC1_BSN	Basin	0.1	1	1/day	Rate constant for oxidation of NH ₃	Yes	0.477	0.981	0.691	0.613	0.992		0.382	
18	RSDCO	Basin	0.02	0.1		Residue decomposition coefficient	Yes	0.099	0.091	0.028	0.036	0.053		0.055	
19	ALPHA_BF	HRU	-100	100	% [€]	Baseflow alpha factor (days)		0.136	-0.416	0.394	-0.341	1.429	-0.476	0.614	-0.331
20	GW_DELAY	HRU	-200	200	%	Groundwater delay (days)		-1.409	-1.264	-1.700	-1.686	-1.825	-1.227	-1.809	-1.551
21	GW_REVAP	HRU	-200	200	%	Groundwater revap coefficient		0.508	-0.369	-1.644	0.393	1.065	0.611	-0.524	0.079
22	GWHT	HRU	0	25	m	Initial groundwater height		6.951	19.017	5.563	4.571	14.340	22.802	6.062	19.705
23	RCHRG_DP	HRU	-200	200	%	Deep aquifer percolation fraction		1.967	2.000	1.261	1.992	-0.362	-1.888	1.460	1.836
24	REVAPMN	HRU	-100	100	%	Threshold depth for revap to occur (mm)	Yes	-0.651	0.156	-0.894	-0.030	-1.962		-1.958	
25	HLIFE_NGW	HRU	0	200	days	Half-life of nitrogen in groundwater		199.44	199.67	194.42	199.90	139.80	197.07	159.27	199.39
26	SHALLST_N	HRU	50	200	mg/l	Concentration of NO ₃ in groundwater		56.081	95.407	52.451	99.396	55.837	113.35	92.618	99.416
27	SHALLST	HRU	0	5E4	mm	Initial depth of water in the shallow aquifer	Yes	48675	48438	48060	47825	1111.7		49366	
28	LAT_ORGN	HRU	0	200	mg/l	Organic N in the baseflow	Yes	11.480	193.76	0.323	197.22	198.48		193.58	
29	CH_K2	Subbasin	-200	200	%	hydraulic conductivity in main channel		-1.558	1.713	0.718	1.569	1.844	1.299	-0.063	0.078
30	CH_N2	Subbasin	-200	200	%	Manning's n value for the main channel		-1.961	1.757	-0.653	1.941	1.945	1.949	-0.756	1.952
31	CH_K1	Subbasin	-200	200	%	hydraulic conductivity in tributary channel	Yes	-0.081	-0.168	1.638	-1.182	-1.232		-0.266	
32	CH_N1	Subbasin	-200	200	%	Manning's n value for the tributary channels		2.000	1.972	1.998	1.997	1.612	-1.512	1.944	1.844
33	SOL_ORGN	HRU	0	100	mg/kg	Initial organic N concentration in the soil layer		65.463	99.851	96.808	5.740	98.802	99.748	98.839	94.181
34	SOL_NO3()	HRU	0	100	mg/kg	Initial NO ₃ concentration in the soil layer		83.262	59.894	99.260	6.852	1.501	1.877	96.435	96.327
35	OV_N	HRU	-200	200	%	Manning's n value for overland flow	Yes	1.994	1.986	1.957	1.967	1.983		1.997	
36	SLSUBBSN	HRU	-200	200	%	subbasin slope length		1.889	0.684	0.908	0.767	1.045	0.245	1.192	0.745
37	CN2	HRU	-25	0	%	SCS runoff curve number	Yes	-0.141	-0.206	-0.144	-0.210	-0.249		-0.196	
38	SOL_AWC()	HRU	-25	25	%	Available water capacity of the soil layer		0.241	0.234	0.244	0.230	0.021	0.225	0.077	0.237

[€] Parameters with unit of % reflect relative change in original parameter value to preserve spatial heterogeneity; the original value is multiplied by the suggested calibration value.

¥ Column identifies constrained parameters for case 3 and case 4

Table 7: Behavior definition represented by the General Performance Ratings (concept adopted from Moriasi et al. 2007 and modified for daily scale simulations)

Performance Rating	NSE		PBIAS (%)	
	Streamflow	NO ₃	Streamflow	NO ₃
Very Good	$0.65 < \text{NSE} \leq 1.00$	$0.55 < \text{NSE} \leq 1.00$	$\text{PBIAS} < \pm 10$	$\text{PBIAS} < \pm 25$
Good	$0.55 < \text{NSE} \leq 0.65$	$0.45 < \text{NSE} \leq 0.55$	$\pm 10 \leq \text{PBIAS} < \pm 15$	$\pm 25 \leq \text{PBIAS} < \pm 40$
Satisfactory	$0.45 < \text{NSE} \leq 0.55$	$0.30 < \text{NSE} \leq 0.45$	$\pm 15 \leq \text{PBIAS} < \pm 25$	$\pm 40 \leq \text{PBIAS} < \pm 70$
Unsatisfactory	$\text{NSE} \leq 0.45$	$\text{NSE} \leq 0.3$	$\text{PBIAS} \geq \pm 25$	$\text{PBIAS} \geq \pm 70$

NSE: Nash-Sutcliffe efficiency coefficient

PBIAS: Percent bias

Table 8: Model performance statistics for various calibration strategies

		Flow			NO ₃		
		NSE	PBIAS	R ²	NSE	PBIAS	R ²
Case 1	Greensboro	0.66	5%	0.66	0.48	14%	0.51
	Tuckahoe	0.59	-17%	0.63	0.34	20%	0.46
Case 2	Greensboro	0.66	1%	0.66	0.47	15%	0.51
	Tuckahoe	0.57	-19%	0.62	0.33	18%	0.43
Case 3	Greensboro	0.65	0%	0.66	0.60	14%	0.63
	Tuckahoe	0.57	-20%	0.61	0.33	25%	0.44
Case 4	Greensboro	0.66	3%	0.67	0.35	16%	0.42
	Tuckahoe	0.56	-15%	0.60	0.31	18%	0.42

A negative PBIAS indicates overestimation

Table 9: NO₃ related internal watershed balances in Greensboro and Tuckahoe models (denoted by G and T, respectively) for different case scenarios. The balances were estimated for agricultural HRUs only. All units are (kg N ha⁻¹ yr⁻¹).

	Case 1		Case 2		Case 3		Case 4	
	G	T	G	T	G	T	G	T
NO ₃ yield from groundwater	8.2	19.7	7.9	19.6	6.8	19.0	8.1	20.3
NO ₃ lost to denitrification	48.5	19.2	42.8	40.5	2.4	2.1	36.7	33.8

Table 10: Autocalibrated rates of denitrification coefficient (β) for soils with different drainage classes in cases 2 and 4

Soil drainage class	Min	Max	Case 2	Case 4
Excessively well drained	0	0.5	0.001	0.412
Well drained	0.5	1	0.513	0.792
Moderately well drained	1	1.5	1.451	1.003
Somewhat poorly drained	1.5	2	1.680	1.618
Poorly drained	2	2.5	2.002	2.004
Very poorly drained	2.5	3	2.720	2.516

Figure 1: The location of Greensboro and Tuckahoe watersheds within the Choptank River basin on the Delmarva Peninsula in the Mid-Atlantic region of North America. The location of USGS flow and USDA water quality stations have been marked on the map.

Figure 2: Box plots of monthly NO_3 loadings from Greensboro and Tuckahoe watersheds. Data collection was performed by USGS and Maryland Department of Environment (MDE) over 8 years (Jan 2001 – Dec 2008) through grab sampling at the outlet of watersheds. The whiskers display 10th and 90th percentile.

Figure 3: Land use (top) and soil drainage class (bottom) maps for Greensboro and Tuckahoe watersheds. Source: USGS National Land Cover Database (NLCD) and USDA Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO).

Figure 4: Flow chart illustrating the parallel calibration approach in SWAT-PMCT with and without distributed denitrification scheme. χ is the maximum number of function evaluations, set to 3000 in this study. Note that modified SWAT.exe used in case 2 and case 4 accommodates the distributed denitrification scheme.

Figure 5: Flow calibration time series for Greensboro and Tuckahoe models under various calibration strategies. The gaps in the time series represent periods with missing observed data. Note that Y axis is in logarithmic scale.

Figure 6: NO_3 calibration time series for Greensboro and Tuckahoe models under various calibration strategies. The gaps in the time series represent periods with missing observed data.

Figure 7: Comparison of observed flow and model responses in varying probabilities for different case scenarios. The highlighted strips on top panels are magnified (bottom panels) to reveal the divergence at high flows ($P > 95\%$)

Figure 8: Comparison of observed NO_3 loading and model responses in varying probabilities for different case scenarios. The highlighted strips on top panels are magnified (bottom panels) to reveal the divergence at high loadings ($P > 95\%$)

Figure 9: Performance of objective function values versus model iterations for Greensboro and Tuckahoe models (denoted by G and T, respectively) in all case scenarios. Note that for case 3 and case 4, the objective function value used for auto calibration is the composed of Greensboro + Tuckahoe NSE values.

Figure 10: Percentage of behavior solutions that satisfy behavior definition (Table 7) in categories of (i) Very good; (ii) Good; (iii) Satisfactory; and (iv) Unsatisfactory.

Figure 11: Aerial image (top left) and soil map (top right) of an agricultural plot in the Greensboro watershed. The bottom panels show the spatial representation of denitrification loss rates estimated by case 1 (left; control) and case 2 (right; distributed denitrification) models.

Figure 1: The location of Greensboro and Tuckahoe watersheds within the Choptank River basin on the Delmarva Peninsula in the Mid-Atlantic region of North America. The location of USGS flow and USDA water quality stations have been marked on the map.

Figure 2: Box plots of monthly NO₃ loadings from Greensboro and Tuckahoe watersheds. Data collection was performed by USGS and Maryland Department of Environment (MDE) over 8 years (Jan 2001 – Dec 2008) through grab sampling at the outlet of watersheds. The whiskers display 10th and 90th percentile.

Figure 3: Land use (top) and soil drainage class (bottom) maps for Greensboro and Tuckahoe watersheds. Source: USGS National Land Cover Database (NLCD) and USDA Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO).

Figure 4: Flow chart illustrating the parallel calibration approach in SWAT-PMCT with and without distributed denitrification scheme. χ is the maximum number of function evaluations, set to 3000 in this study. Note that modified SWAT.exe used in case 2 and case 4 accommodates the

Figure 5: Flow calibration time series for Greensboro and Tuckahoe models under various calibration strategies. The gaps in the time series represent periods with missing observed data. Note that Y axis is in logarithmic scale.

Figure 6: NO₃ calibration time series for Greensboro and Tuckahoe models under various calibration strategies. The gaps in the time series represent periods with missing observed data.

Figure 7: Comparison of observed flow and model responses in varying probabilities for different case scenarios. The highlighted strips on top panels are magnified (bottom panels) to reveal the divergence at high flows ($P > 95\%$).

Figure 8: Comparison of observed NO_3 loading and model responses in varying probabilities for different case scenarios. The highlighted strips on top panels are magnified (bottom panels) to reveal the divergence at high loadings ($P > 95\%$).

Figure 9: Performance of objective function values versus model iterations for Greensboro and Tuckahoe models (denoted by G and T, respectively) in all case scenarios. Note that for case 3 and case 4, the objective function value used for auto calibration is the composed of Greensboro + Tuckahoe NSE values.

Figure 10: Percentage of behavior solutions that satisfy behavior definition (Table 7) in categories of (i) Very good; (ii) Good; (iii) Satisfactory; and (iv) Unsatisfactory.

Figure 11: Aerial image (top left) and soil map (top right) of an agricultural plot in the Greensboro watershed. The bottom panels show the spatial representation of denitrification loss rates estimated by case 1 (left; control) and case 2 (right; distributed denitrification) models.