Journal of Hydrology 590 (2020) 125220

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

# Water quality prediction using SWAT-ANN coupled approach

Navideh Noori[a,*], Latif Kalin[b], Sabahattin Isik[b]

[a] Institute for Disease Modeling, Building IV, 3150 139th Ave SE, Bellevue, WA 98005, USA
[b] School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849 USA

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief

*Keywords:*
Water quality prediction
Hybrid approach
SWAT
ANN
Unmonitored watershed

ABSTRACT

Efficient and accurate prediction of river water quality is challenging due to the complex hydrological and environmental processes affecting their nature. The challenge is even bigger in unmonitored watersheds. Both process- and data-based approaches are utilized for this purpose, with each having its own strengths and weaknesses. The development of a hybrid model can potentially give robust solutions in this regard. To improve the water quality predictions in unmonitored watersheds, we developed a hybrid model by combining a process-based watershed model and artificial neural network (ANN). Combining these two models helped to optimize the calibration and validation process while accounting for the complex hydrological and water quality processes. The developed model was applied to watersheds in the Atlanta metropolitan area, USA, to predict monthly nitrate, ammonium, and phosphate loads. We treated the watersheds as unmonitored and tested the skill of the hybrid model accordingly. The hybrid model had good skills in predicting all three constituents. The model worked especially well for nitrate. As a matter of fact, it even outperformed SWAT models calibrated at each site. This work emphasizes the potential benefits of the proposed hybrid modeling framework for the prediction of water quality parameters in unmonitored watersheds.

## 1. Introduction

Water quality management enhances the ecosystem and human health and helps sustain drinking water production (Ho et al., 2019). For these purposes, water quality data are needed to identify long-term trends, regional variability, emerging problems, etc. Water quality monitoring and sampling over a continuous time period are costly and time-consuming. This downside limits datasets to sparse sampling points throughout the year and restricts conducting water resource management studies as well as calibrating and validating water quality models (Libera and Sankarasubramanian, 2018).

To overcome this limitation towards better water quality management, the development of water quality models is an important step. Different modeling techniques have been developed over the past decades to improve the prediction accuracy of water quality parameters. Statistical models such as multiple linear regression models (Herrig et al., 2015) and regression trees (Stidson et al., 2012), which have been used for water quality prediction, have certain limitations. Most statistical models assume linear and normally distributed associations among the predictors and response variable (Ahmed et al., 2019), do not incorporate hydrological processes (Tongal and Booij, 2018), and their predictions are restricted to gauged watersheds (Ho et al., 2019). On the other hand, process-based models are powerful tools that can be used to simulate hydrological processes and fate and transport of pollutants under different scenarios. However, these models, such as the Soil and Water Assessment Tool (SWAT) (Neitsch et al., 2011), need a large amount of data and processing time (Ahmed et al., 2019) and involve several parameters that cannot be measured (Abdulmohsin et al., 2016). Besides, these models have a long process of parameter estimation, calibration, and validation.

Machine Learning (ML) based methods, such as Artificial Neural Networks (ANN), are increasingly being used for solving environmental problems. These data-based methods can tackle highly nonlinear problems (Abdulmohsin et al., 2016; Hunter et al., 2018; Barzegar et al., 2016; Adnan et al., 2019a; Adnan et al., 2019b) and do not require knowledge of the physical processes, yet often require large volumes of data. A number of studies have assessed the prediction accuracy of water quality constituents using ANN models (Haghiabi et al., 2018; Kalin et al., 2010; Keshtegar and Heddam, 2018; Khataar et al., 2018; Sarkar and Pandey, 2015; Šiljić Tomić et al., 2018; Sirisha et al., 2008; Zhang et al., 2016). ANN applications have their own challenges, such as determining the appropriate network structure, which is obtained through experience and trial and error, and selecting the best combination of the input variables. In addition, extrapolation beyond the

* Corresponding author.
*E-mail address:* nnoori@idmod.org (N. Noori).

https://doi.org/10.1016/j.jhydrol.2020.125220
Received 30 December 2019; Received in revised form 18 June 2020; Accepted 20 June 2020
Available online 26 June 2020

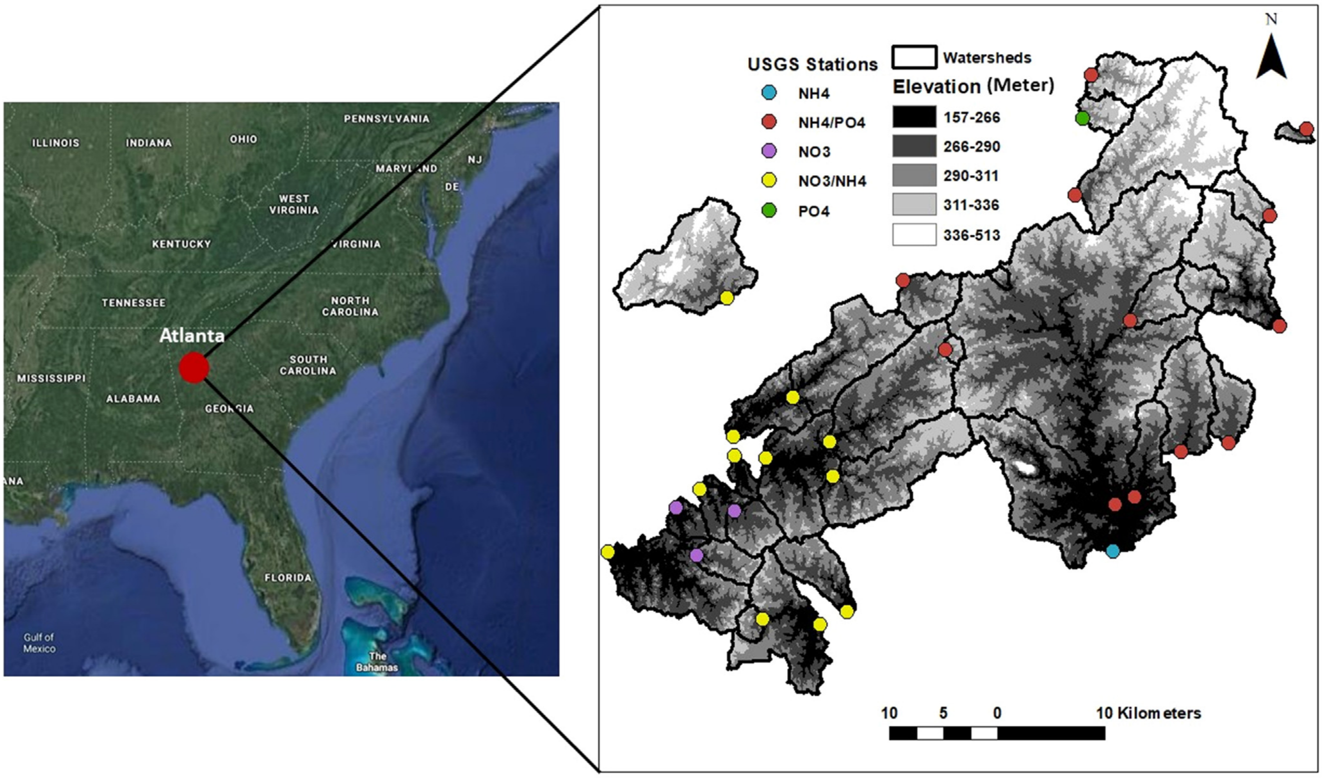


**Fig. 1.** USGS stations with their corresponding available' water quality data around city of Atlanta, Georgia, USA. All the sites have daily streamflow data as well.

range of original training data could lead to large errors. When the land use/cover of watershed changes, ML-based models can fail to capture the impacts of these dynamics on water quality. To develop a more accurate predictive model and account for the limitations of each technique, a hybrid approach could work better than a single model (Rezaeianzadeh et al., 2018; Tongal and Booij, 2018). Previous studies have shown that coupling different techniques can lead to noticeable improvements in water quantity and quality predictions. Examples of such hybrid approaches are developing process-driven ANN and regression models (Hunter et al., 2018), wavelet-ANN model (Alizadeh and Kavianpour, 2015; Barzegar et al., 2016), machine learning methods coupled with base flow separation (Rezaeianzadeh et al., 2018; Tongal and Booij, 2018), adaptive neuro-fuzzy models (Yaseen and Ramal, 2018), and ANN coupled with SCS Curve Number method (Isik et al., 2013). Previous studies have not combined an ML method and a process-based watershed model to predict water quality parameters, especially in unmonitored watersheds. In this study, we extend the hybrid approach developed by Noori and Kalin (2016) for daily streamflow predictions to water quality prediction in unmonitored watersheds. In Noori and Kalin (2016), ANN and SWAT were coupled to overcome the limitations of each model. The coupled model resulted in significant improvements in daily streamflow predictions. Here, we illustrate the application of the SWAT-ANN approach for water quality prediction and test the model performance by applying it to the data from watersheds in and around the Atlanta Metropolitan area, USA.

## 2. Methods

### 2.1. Soil water assessment tool (SWAT)

SWAT is a watershed-scale, semi-distributed, and continuous-time hydrologic and water quality model. SWAT is designed to simulate discharge, sediments, nutrient and pesticide loads, crop growth, and management practices over long time periods at daily and sub-daily scales (Neitsch et al., 2011). The model divides the watersheds into sub-watersheds, which are further subdivided into hydrological response units (HRUs). Each HRU has a unique combination of soil type, land use/cover, and slope. SWAT has been used widely in the field of water resource management for water quantity and quality prediction as well as scenario testing such as effects of land use/land cover changes and climate changes on hydrologic cycle (Bauwe et al., 2019; Chotpantarat and Boonkaewwan, 2018; Kavian et al., 2018; Malagó et al., 2017; Shi et al., 2017; Wu et al., 2017; Yan et al., 2019).

SWAT simulates the hydrologic cycle in two phases of land and water. The land phase controls the amount of water, sediment, and nutrients loadings into the main channel of each sub-watershed and the water phase is defined as the process of routing runoff, sediment, and nutrients through the stream network to the watershed outlet (Neitsch et al., 2011). Nitrogen is simulated by SWAT in the soil profile, taking into consideration five different organic and inorganic pools. Nitrate ($NO_3^-$) and Ammonium ($NH_4^+$) are inorganic forms of nitrogen. Phosphorus is also simulated by SWAT in the soil by monitoring six different organic and inorganic pools. A detailed model description can be found in Neitsch et al. (2011).

For this work, we used the SWAT model developed in the previous study by Noori and Kalin (2016). In addition to the concentrations and loads of various constituents in the stream segments (called reach in SWAT) and watershed outlet, SWAT also provides users partitioning of sources from the sub-watersheds. The amount of monthly $NO_3^-$ loads transported by surface runoff, lateral flow, and baseflow to streamflow, as well as simulated monthly streamflow at the watershed outlet were used as inputs to the $NO_3^-$ ANN model. For $NH_4^+$, SWAT estimated monthly $NH_4^+$ loads at the watershed outlet as well as monthly streamflows were used as inputs to the $NH_4^+$ ANN model. For phosphate ($PO_4^{3-}$), SWAT simulated monthly mineral phosphorus loadings at the watershed outlet, and simulated streamflows were used as inputs to the $PO_4^{3-}$ ANN model.

## 3. Study area and data

The study area is in the southeast of the United States, the state of Georgia, Atlanta. Atlanta and its metropolitan area have a warm and humid climate with an average annual temperature of 16.5 $^{0}$C and average annual precipitation of 1200 mm. Urban development has been very rapid over the past decades in this region, and based on the 2006 National Land Cover Data (NLCD), the dominant land use types are impervious surfaces as well as evergreen/deciduous/mixed forest.

The input data to the model include daily air temperature data from the National Climatic Data Center stations in Atlanta area and daily precipitation data from the North American Land Data Assimilation System (NLDAS), land use/cover data from the 2006 National Land Cover Data (NLCD), and soil data from the Soil Survey Geographic Database (SSURGO). We considered 29 US Geological Survey (USGS) monitoring stations nearby the city of Atlanta with daily streamflow data as well as water quality data available over the period 2002–2010. The locations of these stations with available water quality data are shown in Fig. 1. The summary of available data for each station is also given in Supplementary Material, Table 1. Watersheds ranged in size from 3 to 552 km$^2$. Dominant land cover/use types in these watersheds are impervious surfaces ranging from 13% to 52%, and deciduous forests ranging from 2% to 24%. Overall, 15 stations with $NO_3^-$ data, 25 stations with $NH_4^+$ data and 13 stations with $PO_4^{3-}$ data were used in the study. $NO_3^-$ concentrations ranged between 0.18 and 13 mg/L. Station 2,203,700 had the highest $NO_3^-$ concentration among 15 stations, about two times higher than other stations. $NH_4^+$ concentrations ranged from 0.01 to 10 mg/L. Station 2,334,885 had the highest $NH_4^+$ concentration. Station 2,336,240 overall had very low $NH_4^+$ concentrations, which ranged between 0.03 and 0.16 mg/L. $PO_4^{3-}$ concentrations ranged from 0 to 1.32 mg/L, and station 2,334,480 had the highest concentration among 13 stations.

This study focused on predicting nutrient loads, rather than nutrient concentrations. Accurate prediction of nutrient loads is important for water bodies. Too much nitrogen and phosphorous in the water can lead to various problems, such as algal blooms. Not only can some algae produce toxins, but also, when algae die, they can lead to reductions in oxygen levels in the water. Since the observed water quality concentration data are instantaneous and not in a continuous time scale, the LOAD ESTimator (LOADEST) tool developed by USGS (Runkel et al., 2004) was used to generate continuous time series of each constituent loads at each USGS station. This tool combines streamflow time series with instantaneous water quality concentrations and uses a set of regression models to estimate mean loads over a specified time interval on a continuous scale (daily, monthly, or seasonal). This tool has been used commonly to generate continuous data from discrete data and to estimate monthly or annual loading in SWAT modeling (Dagnew et al., 2016; Lee et al., 2018; Niraula et al., 2013; Singh et al., 2015: Wallace et al., 2018; Wang et al., 2016). The LOADEST performance ($R^2$) ranged from 0.84 to 0.96 for $NO_3^-$, between 0.53 and 0.95 for $NH_4^+$, and from 0.46 to 0.83 for $PO_4^{3-}$.

## 4. Water quality prediction

### 4.1. Coupling SWAT and ANN

Using the SWAT models from the previous study (Noori and Kalin, 2016) for each station and setting the model parameters at their default values, monthly water quality loads were simulated. SWAT generated outputs were then used as inputs to the ANN model. SWAT was intentionally not calibrated because our goal was to develop a model that can be used in unmonitored watersheds. For each constituent, a separate ANN model was developed. The continuous monthly water quality loads, estimated by LOADEST, served as the observed data to train the ANN models.

ANN is inspired by the structure of the human brain and has the ability to model complex nonlinear relationships, without the detailed knowledge of the internal functions of a system (Kalin et al., 2010). ANN is classified based on the number of layers and the direction of information flow. Feed-forward network with Levenberg–Marquardt back-propagation learning has been successfully applied to hydrological and environmental problems (Kalin et al., 2010). The feed-forward network includes three layers of input, hidden and output. These layers are connected with each other through neurons. Input and output layers have neurons equal to the number of inputs and outputs. For the hidden layers, the optimum number of layers and the number of neurons within hidden layers are usually found through the trial-and-error approach. In this study, the tangent sigmoid transfer function was adopted for both hidden and output layers, and the sum of square error was minimized as the error function. For the model training and testing purposes, the leave-one- site-out jackknifing technique was used (Sefick et al., 2015). In this technique, one observation is left out of the training data set, the model is retrained, and the observation that was left out is predicted. In this study, out of $n$ stations, one was left for testing, and the model was trained using the remaining ($n$-1) stations data. This step was repeated until all stations had been removed once. This technique was explained in Noori and Kalin (2016) in more detail. The schematic of the coupled model is given in Fig. 2. The performances of the models were evaluated with the coefficient of determination ($R^2$), Nash–Sutcliffe efficiency ($E_{NASH}$) (Nash and Sutcliffe, 1970) and percent bias ratio ($P_{BIAS}$) (Salas et al., 2000).

The coefficient of determination is a measure of linear correlation between two quantities and is given by:

$$R^2 = \left( \frac{n\sum O_i S_i - \sum O_i \sum S_i}{\sqrt{n(\sum O_i^2) - (\sum O_i)^2} - \sqrt{n(\sum S_i^2) - (\sum S_i)^2}} \right)^2$$

where, $O$ and $S$ represent observed data and model outputs, and $n$ is the number of data points. The Nash–Sutcliff efficiency ($E_{NASH}$) is commonly used to assess the predictive power of hydrological models (Nash and Sutcliffe, 1970). It is defined as:

$$E_{NASH} = 1 - \frac{\sum (O_i - S_i)^2}{\sum (O_i - \bar{O})^2}$$

where, $\bar{O}$ is the mean of the observed data. The percent bias ratio is expressed as

$$P_{BIAS} = 100\frac{\sum (S_i - O_i)}{\sum O_i}$$

The bias ratio measures the degree to which the forecast is under- or overpredicted. A negative bias ratio indicates underprediction, whereas a positive bias ratio reflects overprediction (Kalin et al., 2010).

The developed ANN models performances were rated using the model evaluation guidelines developed by Moriasi et al. (2007) for nutrients based on $P_{BIAS}$ and $E_{NASH}$:

Very good: $0.75 < E_{NASH} \le 1$; $|P_{BIAS}| < 25$
Good: $0.65 < E_{NASH} \le 0.75$; $25 \le |P_{BIAS}| < 40$
Satisfactory: $0.50 < E_{NASH} \le 0.65$; $40 \le |P_{BIAS}| < 70$
Unsatisfactory: $E_{NASH} \le 0.50$; $|P_{BIAS}| \ge 70$

The architecture of the neural network utilized in this study is shown in Fig. 3. The proposed feed-forward neural network has three main layers: input, hidden, and output layers. The input layer has observed streamflow and constituent load obtained from the SWAT model. The hidden layer has multiple neurons. The number of neurons in the hidden layer varies with water quality types and sites. Determination of the optimum number of layers is usually a matter of experimentation. A trial-and-error approach is the most commonly used method to find the number of hidden neurons and layers (Kalin et al., 2010). In this study, the number of hidden neurons was searched from 1

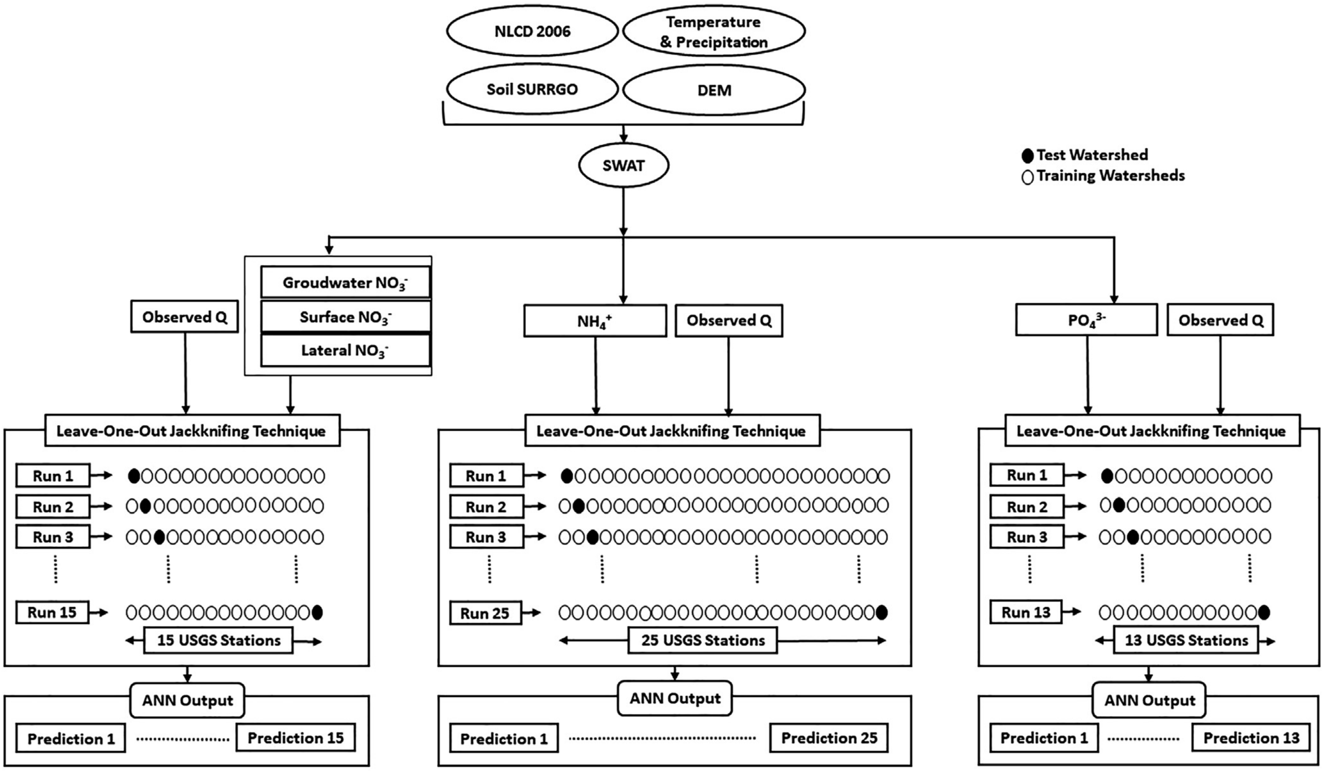


**Fig. 2.** Schematic of coupled SWAT-ANN model for nitrate, ammonium and phosphate loads. Q is streamflow.

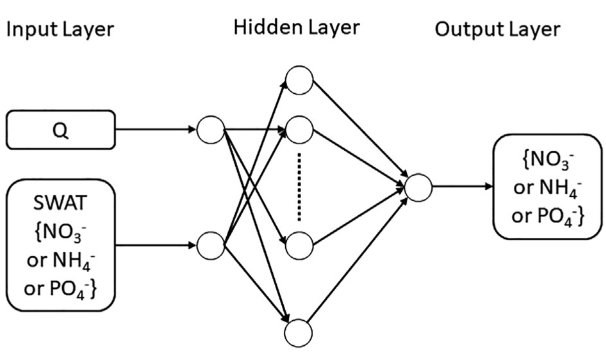


**Fig. 3.** The architecture of neural network.

to 30, respectively. The output layer is the predicted constituent i.e., $NH_4^+$, $NO_3^-$ or $PO_4^{3-}$.

Mean square error (*MSE*) and Akaike's information criterion (*AIC*) are used as selection criteria in determining optimal input and hidden neurons. The *AIC* is commonly used in the literature to find optimal ANN architectures (Kalin et al., 2010; Qi and Zhang, 2001). Various forms of *AIC* are used in the literature. We used the one proposed by Qi and Zhang (2001):

$AIC = \log(\sigma^2_{MLE}) + 2m/n$ if $n/(m+1) \geq 40$

$AIC = \log(\sigma^2_{MLE}) + 2m/(n-m-1)$ if $n/(m+1) < 40$ where, $n$ is the number of data, and $m$ is the number of parameters in the model. The term $\sigma^2_{MLE}$ denotes the maximum likelihood estimate of the variance of the residual term or simply the mean square error (*MSE*) between the observed and simulated data. MATLAB version 9.2.0 (2017a) was used for ANN model simulations.

To show the advantage of the SWAT-ANN model over the traditional ANN model, we developed ANN models similar to ones in Kalin et al. (2010) for $NO_3^-$, $NH_4^+$, and $PO_4^{3-}$ using LULC, streamflow, and temperature as input variables. Again, the jackknifing method was used for training and testing the models. The performance of the ANN model was compared with the hybrid model.

## 5. Swat-CUP

To have another level of assessment for the prediction power of the SWAT-ANN model, we compared the performance of the coupled model with SWAT by calibrating the SWAT model for $NO_3^-$, as a check against possible idiosyncrasies of the hybrid approach. We did not calibrate the SWAT model for $PO_4^{3-}$ and $NH_4^+$, considering the main goal, which solely was for comparison.

The SWAT model was calibrated using the SWAT Calibration Uncertainty Procedure (SWAT-CUP) (Abbaspour et al., 2007; Abbaspour, 2015). SWAT-CUP is a calibration, sensitivity, and uncertainty analysis tool for SWAT. In this study, the Sequential Uncertainty Fitting (SUFI-2) module (Abbaspour et al., 2007) was used for model calibration. SUFI-2 determines the most sensitive parameters for calibration through Latin Hypercube Sampling. SWAT-CUP was run for each station separately on the monthly time step. The number of iterations varied from 3 to 10. Each iteration had 500 simulations. For the first iteration, the parameters that were considered for the daily streamflow calibration in the previous study (Noori and Kalin, 2016) were fixed to their best-estimated values, then the monthly $NO_3^-$ load was added as observed data to SWAT-CUP and new set of parameters were added for $NO_3$calibration with their default ranges. For the next iterations, SWAT-CUP recommended parameter ranges were used. When the SWAT-CUP recommended range of a parameter exceeded its min/max limit, the recommended range was adjusted. Model performance was evaluated with $R^2$, $E_{NASH}$, and $P_{BIAS}$.

## 6. Results

The testing results of SWAT-ANN hybrid models are given in Fig. 4. Figs. 5–7 compare the observed monthly load time series with the model simulations. It is necessary to emphasize one more time that each watershed in these figures was treated as an unmonitored watershed for

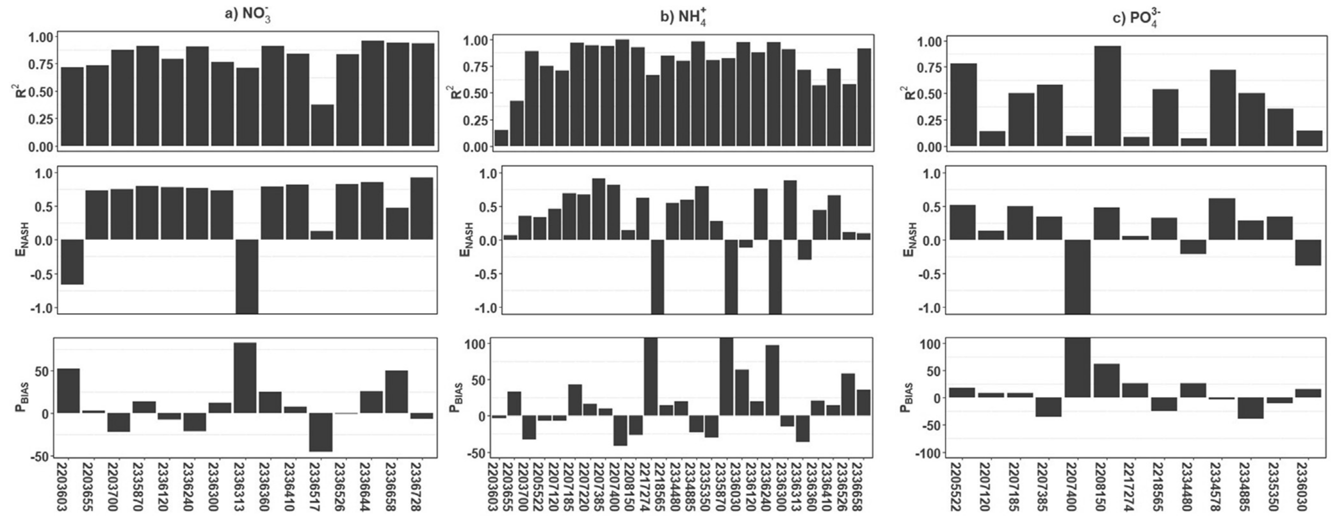


**Fig. 4.** Testing results of SWAT-ANN model for a) $NO_3^-$ load at 15 stations, b) $NH_4^+$ load at 25 stations, c) $PO_4^{3-}$ load at 13 stations.

testing purposes. When a watershed was used for testing, no data from that watershed was used at all during the training of the SWAT-ANN model. For the stations with $NO_3^-$ data, the training performance rating was "very good" with $E_{NASH} > 0.80$ for all the stations (Supplementary Material, Table 2). For the testing, more than 70% of the watersheds had the performance rating of "good" to "very good". The median $E_{NASH}$ and absolute $P_{BIAS}$ values for $NO_3^-$ runs were 0.77 and 7%, respectively. If the two sites with very low performances are ignored, these values improve to 0.78 and 3%, respectively. As shown in Fig. 5, SWAT-ANN simulated $NO_3^-$ load successfully captured the observed time series pattern and peaks and troughs for the majority of the 15 stations. The hybrid model overestimated the $NO_3^-$ load in stations 2,203,603 and 2,336,313 which have small watershed areas, 6 and 4 km$^2$ respectively, with impervious surface and urban grass as the dominant land use/land cover types. The majority of watersheds used for training of the $NO_3^-$ models had area > 10 km$^2$. The model did not capture the peak load at the beginning of the time series in station 2,336,517 but captured the observed pattern later in the time series (Fig. 5). Plotting the $E_{NASH}$ values against the watershed characteristics (Fig. 8a) showed that as the watershed size increases, the model performance rate increases as well (*p-value* = 0.003). Also, as the percent impervious cover increases in these watersheds, the model performance rate decreases (*p-value* < 0.001). On the other hand, watersheds with a higher percentage of forest cover had better model performances (*p-value* = 0.03). We found a correlation between watershed area and level of imperviousness or forest cover in the study region ($r = -0.15$ and 0.16, respectively), although there is a significant correlation between forest cover and imperviousness ($r = -0.91$). Therefore, the opposite trends between the model performance versus percent impervious cover and the model performance versus watershed size are not related. One can speculate that as the imperviousness (thus urban land) increases, the likelihood of having point or any other nutrient sources, such as combined sewer overflows (cso) or leaky sewer systems increases. Our modeling framework did not consider such sources. These results highlight that the developed SWAT-ANN model predicts the $NO_3^-$ load in forested watersheds (or less urbanized) with the area more than 10 km$^2$ in the Atlanta metropolitan area with higher accuracy.

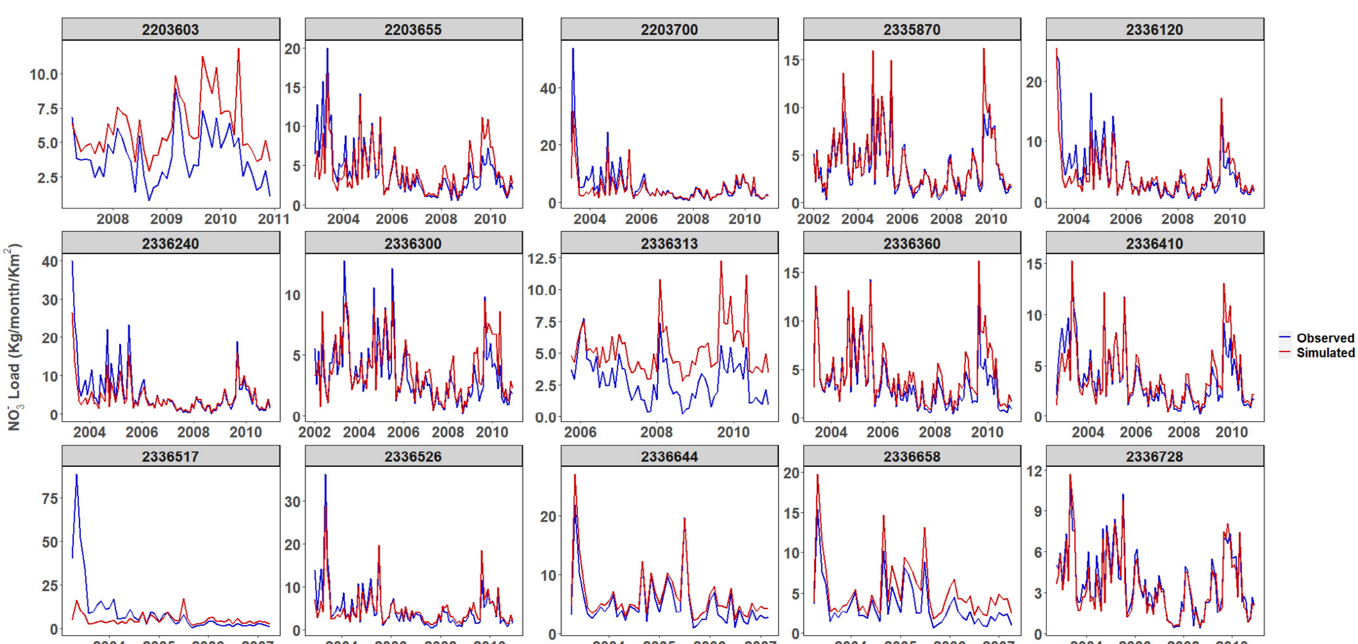


**Fig. 5.** SWAT-ANN simulated $NO_3^-$ load time series versus observed data for 15 stations.

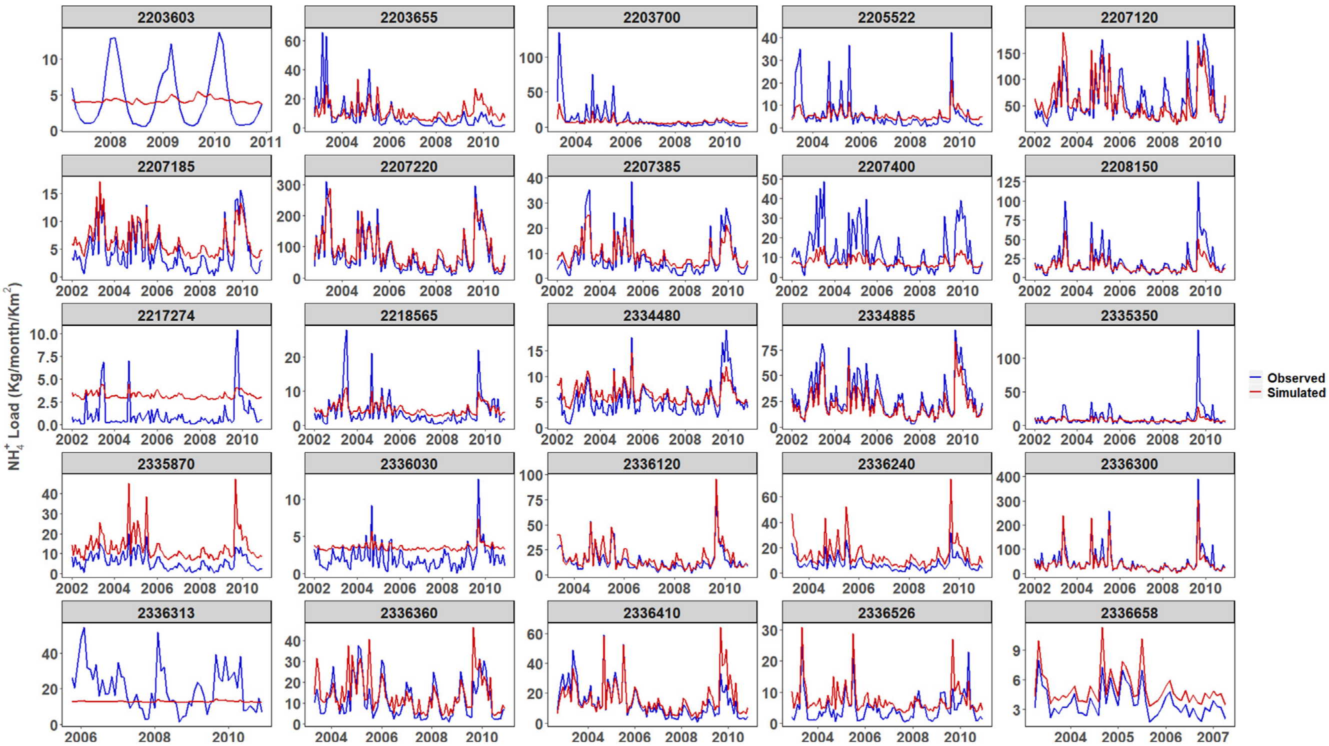


**Fig. 6.** SWAT-ANN simulated $NH_4^+$ load time series versus observed data for 25 stations.

For $NH_4$, the training performance rating was "very good" with $E_{NASH} > 0.75$ for 24 out of the 25 stations (Supplementary Material, Table 2). For the testing, 11 out of 25 stations had $E_{NASH} > 0.50$, and the performance rating was "satisfactory" to "very good", with the median $E_{NASH}$ and absolute $P_{BIAS}$ values at 0.44 and 15%, respectively. The SWAT-ANN model could not capture the large peaks of the $NH_4$ load for the majority of stations but predicted the small peaks more accurately. The models, except for stations 2,336,313 and 2203603, also predicted the variations in the data accurately. $NO_3^-$ load at these two stations was also overestimated by the hybrid model. For some of the stations, the model underestimated the observed peak loads and overestimated the troughs (Fig. 6). The majority (88%) of watersheds with $NH_4^+$ data in this study have less than 37% impervious cover (median = 23%). However, the watersheds draining to stations 2203603, 2336030, and 2,336,313 have an impervious cover of 46%, 52%, and 51%, respectively, and are the most urbanized watersheds (Supplementary Material, Table 1). These three stations also have much higher observed concentrations (mean is 1.31 mg/L) than the other stations (mean 0.09 mg/L). Therefore, with the leave-one-out jack-knifing, the trained models are having a hard time predicting $NH_4^+$ levels at these sites. Linking the area of land use/land cover characteristics of these watersheds with the $NH_4^+$ model performance rate did not add any strong explanation to the analysis. However, if we kick out the two watersheds with very low performances, then we observe

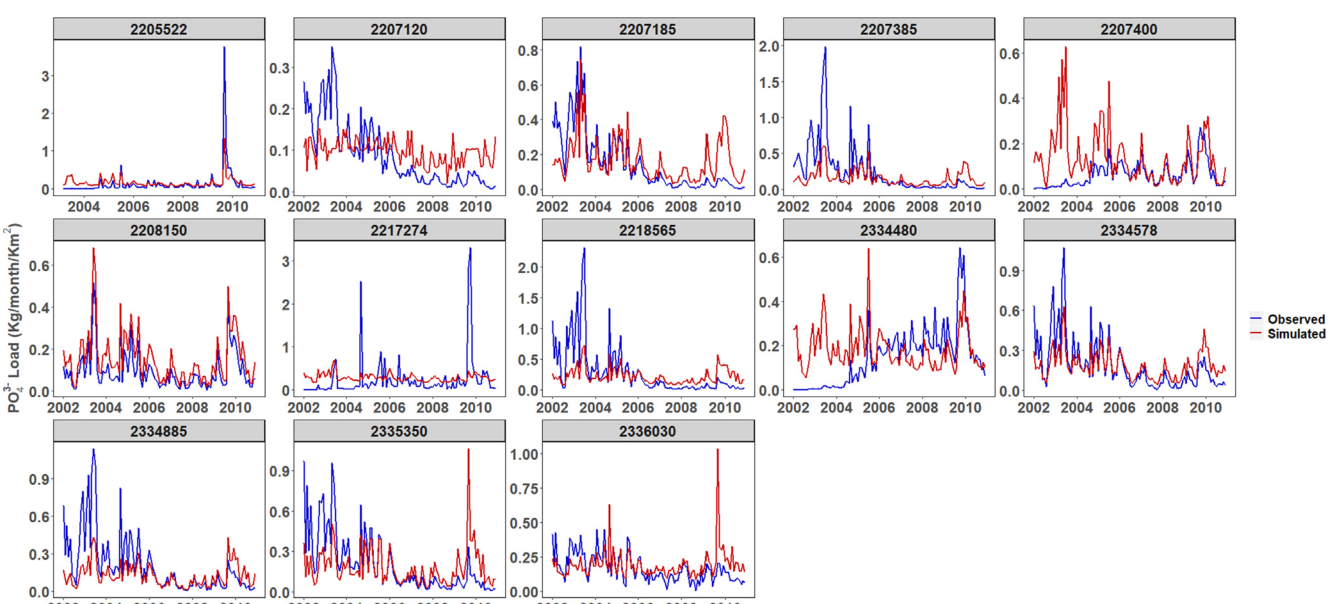


**Fig. 7.** SWAT-ANN simulated $PO_4^{3-}$ load time series versus observed data for 13 stations.

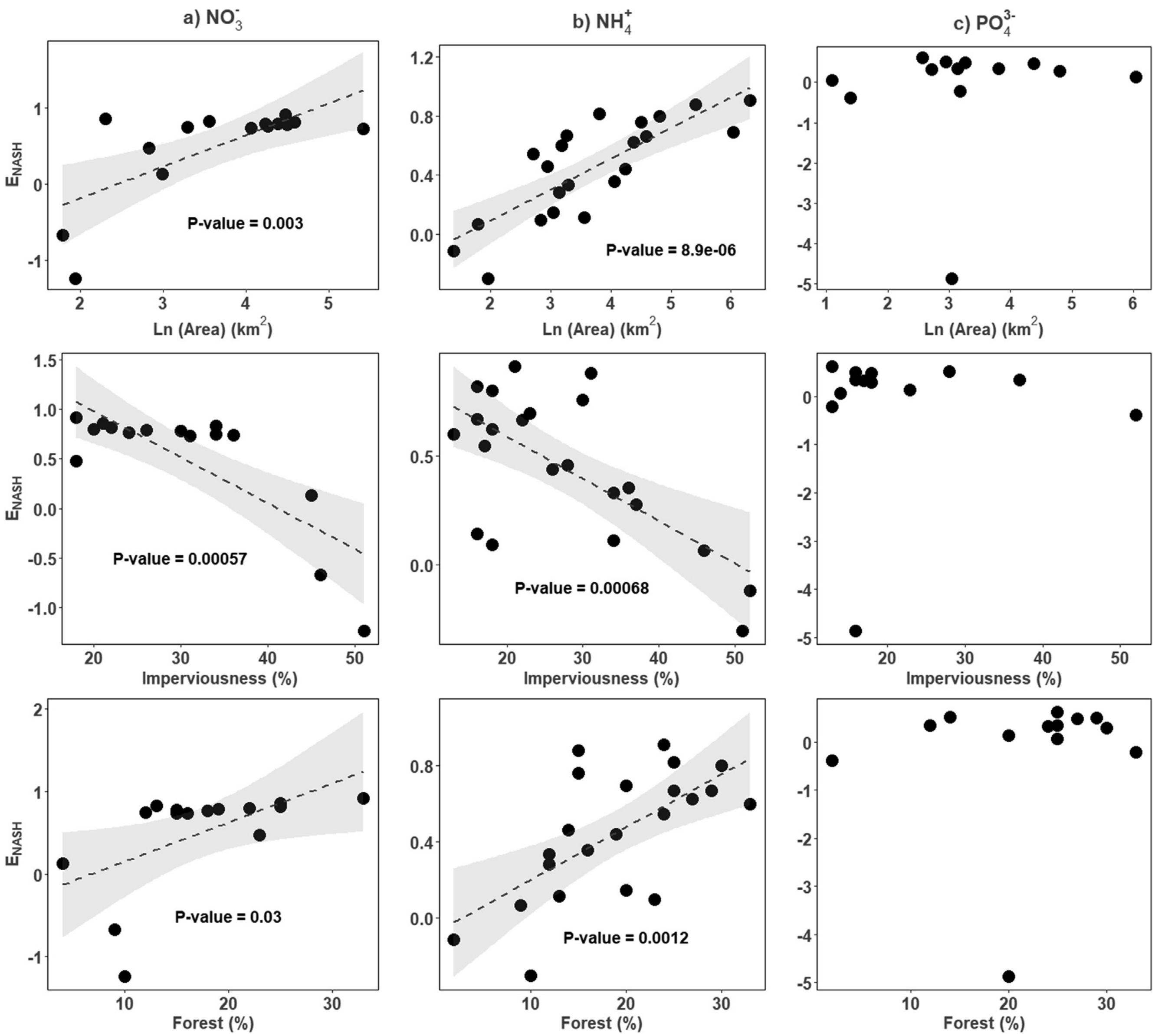


**Fig. 8.** $E_{NASH}$ values of SWAT-ANN model for a) $NO_3^-$, b) $NH_4^+$ and c) $PO_4^{3-}$ versus test watersheds natural logarithm of area, percent imperviousness and percent forest cover.

the same trends we observed with $NO_3^-$ (Fig. 8b). Station 2,217,274 has the smallest watershed area at 3 km$^2$, (for the rest, median = 35 km$^2$, mean = 85 km$^2$).

For $PO_4^{3-}$, the training performance rating was between "satisfactory" to "good" with $E_{NASH} > 0.55$ for all the stations (Supplementary Material, Table 2). For the testing, the developed model of $PO_4^{3-}$ load, similar to $NH_4^+$, underestimated the large peaks (Fig. 7) and had the weakest performance ratings with only 4 out of 13 stations with $E_{NASH}$ value of about 0.50 or higher (Fig. 4c). There were no strong associations between either of the watershed area, percent imperviousness or forest cover, and the model performance rate (Fig. 8c). Removing the watersheds with very low performances did not change results, and no statistically significant associations were found. However, the percent of pasture cover was significantly related to the hybrid model performance rate (*p-value* = 0.023). Watersheds with higher pasture cover had the lowest $E_{NASH}$ values. This result highlights that the source of phosphorus from the pasture was not accounted for and was not simulated well by the SWAT model. Note that we used the default SWAT parameterization. No adjustments were made to any of the SWAT parameters.

## 7. Discussion

For comparison purposes, we calibrated the SWAT model using SWAT-CUP for monthly $NO_3$ loads. The SWAT-CUP performance rating was "unsatisfactory" for all 15 stations with $NO_3^-$ data (Fig. 9). The mean $E_{NASH}$ and absolute $P_{BIAS}$ values for $NO_3^-$ runs were 0.06 and 6%, respectively. The testing step of SWAT-CUP includes selecting the nearest watershed to the trained watershed and running the SWAT-CUP model with the trained model's best-estimated parameters. However, considering the weak performance of SWAT-CUP for the training set, we chose to skip the testing step. Comparing the SWAT-CUP calibration results with the hybrid model testing results, it is clear that the SWAT-ANN model outperformed the SWAT model, and coupling SWAT and ANN improved the $NO_3^-$ prediction accuracy.

We also compared the predictive power of ANN models developed using LULC, temperature, and streamflow as input variables with the hybrid model. Based on the $E_{NASH}$ values, for $NO_3^-$, the hybrid model

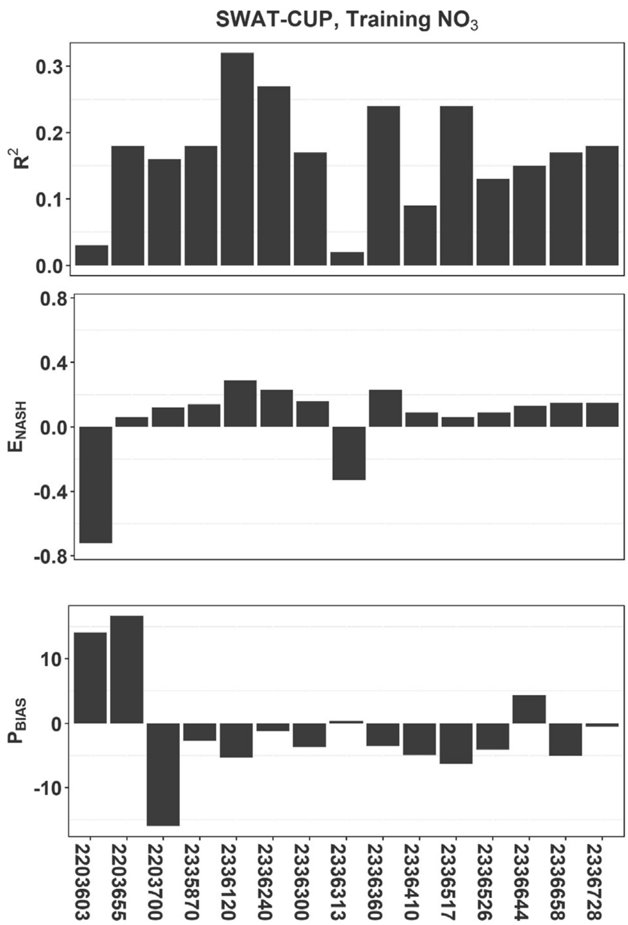


**Fig. 9.** Model performance of SWAT for $NO_3^-$ load at 15 stations after calibration with SWAT-CUP.

outperformed the ANN model in 85% of the stations. The mean/median $E_{NASH}$ values were 0.50/0.77 and −4.3/0.73, respectively, for SWAT-ANN and ANN. For $NH_4^+$, the hybrid model had a better performance than the ANN model in 56% of the stations. The mean/median $E_{NASH}$ values were 0.15/0.44 and −1.2/0.48, respectively, for SWAT-ANN and ANN. For $PO_4^{3-}$, in 77% of the stations, SWAT-ANN performed better. The mean/median $E_{NASH}$ values were −0.14/0.32 and −0.32/0.01, respectively, for SWAT-ANN and ANN. Based on performance measures, this comparison highlights the superiority of the coupled model.

Comparing the results of this study with our previous study (Noori and Kalin, 2016) revealed reverse trends between streamflow model performance and watershed characteristics. Although many of the stations used in this study overlap with the ones in Noori and Kalin (2016), they are not exactly the same. In Noori and Kalin (2016), there were 29 stations. On the other hand, in this study, we had 25 sites for $NH_4$, only 15 sites for $NO_3^-$ and 13 sites for $PO_4^{3-}$. Further, the objective in Noori and Kalin (2016) was predicting daily streamflow. In this study, the focus was on predicting monthly nutrient loads. The processes and time scales of the two studies are completely different. In addition, Noori and Kalin (2016) had separate models for cool and warm seasons to account for seasonal variations in contribution of baseflow and stormflow to the total flow as well as to account for environmental and vegetation variations. Considering all these factors, it is hard to make a comparison between the two studies.

The SWAT-ANN training performance somewhat mirrors the LOADEST performance. $PO_4^{3-}$ model performance is lower than $NO_3^-$ and $NH_4^+$ models because the phosphorus cycle is more complex; it is simulated by SWAT by monitoring six different organic and inorganic pools. In addition, some of the phosphorous are transported to the main channels with sediment. If we had sediment data, $PO_4$ predictions would have been likely better, both for LOADEST and SWAT-ANN. The LOADEST performance for $NO_3^-$ and $NH_4^+$ are not too far apart, with the former being a little better. The reason for the slightly lower performance of $NH_4^+$ is likely due to the higher variation in $NH_4^+$ measurements. The average coefficient of variation for observed $NO_3^-$ concentrations was 0.38 (0.29–0.50). For observed $NH_4^+$ concentrations, it was 1.17 (0.45–4.2). The SWAT-ANN model performances for $NO_3^-$ and $NH_4^+$ during training are also close. Note that there is no training/testing sequence with LOADEST. Therefore, comparing SWAT-ANN testing performance to LOADEST is not meaningful.

## 8. Summary and conclusions

This work expanded a hybrid-modeling framework, previously developed for streamflow prediction in ungauged watersheds by Noori and Kain (2016), to monthly nutrient load prediction in unmonitored watersheds. Testing of the model for predicting $NO_3^-$, $NH_4^+$, and $PO_4^{3-}$ in the Atlanta metropolitan area, Southeast USA, showed that the hybrid model outperformed standalone SWAT and ANN models. The developed hybrid model improved the water quality prediction accuracy by accounting for the hydrology as well as fate and transport of nutrients on land and streams through incorporation of the SWAT model into ANN.

If reliable water quality data is available, a calibrated and validated SWAT model can be a robust tool for assessing the impacts of land-use change, climate change/variability, or any management operations in a watershed. If water quality data is not available, then SWAT is typically calibrated and validated at a nearby watershed with similar characteristics, and the model parameters are transferred from the donor watershed to the target watershed (Wang and Kalin, 2011). The hybrid model developed in this study eliminates the need for parameter transferring. In our study, the hybrid model had better skills in predicting nutrient loads than the SWAT model calibrated at each individual site.

A comparison of the hybrid model to the ANN model showed that the prediction accuracy of the hybrid model was higher for the majority of watersheds for all three nutrients. The input variables to the ANN model were LULC percentages, temperature, and streamflow. In the hybrid model, LULC and temperature were replaced with SWAT-simulated nutrient loads, which helped us better represent the hydrological and water quality processes. SWAT captures effects of LULC on hydrology through various processes, such as canopy interception, evapotranspiration (ET), plant growth, runoff generation, and infiltration (CN varies by LULC in the SCS-CN method), and overland flow routing (through Manning's roughness). Similarly, SWAT simulates movements of nutrients from land to the stream network and calculates the nutrient loadings for each HRU in the watershed, which increases the accuracy of the loading prediction from the watershed. SWAT also accounts for the loadings from point sources, such as the wastewater treatment plant, and atmospheric deposition. Best Management Practices (BMPs), implemented for nutrient and sediment load reduction, can easily be represented in the SWAT model. Capturing such detailed information is not possible with ANN.

The hybrid model developed for monthly $NO_3^-$ load outperformed $NH_4^+$ and $PO_4^{3-}$ hybrid models substantially. This could be due to including the SWAT-simulated lateral flow, surface flow, and groundwater $NO_3^-$ contributions to streamflow as inputs to the ANN model. Overall, the hybrid model could capture the observed variations in all three nutrients load; however, it underestimated the large peaks of $NH_4^+$ and $PO_4^{3-}$ data. The results of this study highlight the potential benefits of the proposed framework for water quality components prediction. The hybrid model can be considered as a regionalized approach in an area with unmonitored watersheds where there are water quality and flow data available in neighboring watersheds with similar landscape characteristics. Application of our hybrid model in such regions leads to optimization of the water quality monitoring by lowering the cost and process time.

Last but not least, although the developed hybrid models can be reliably implemented in watersheds with similar characteristics in and near the Atlanta metropolitan area, the methodology is universal and can be developed in any other region in the world with low landscape heterogeneity, since extrapolation beyond the range of original training data would lower the model accuracy. Similar hybrid models can be developed using other watershed models, such as HSPF, SWMM, MIKE-SHE, etc., and machine learning techniques (random forest, gradient boosting machine, extreme learning machine, M5-cubist, elastic net, etc.). The idea of coupling process-based models and machine learning techniques, with the goal of optimizing the computational process and increasing the prediction accuracy, can be extended beyond predicting streamflow and stream water quality to solve many other complex problems.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This project was funded by the USDA Forest Service, National Urban Community Forestry Council, United States, and the Center for Environmental Studies at the Urban-Rural Interface, Auburn University, United States.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125220.

## References

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

Abbaspour KC, 2015 SWAT-CUP SWAT Calibration and Uncertainty Programs-A User Manual. Tech. rep., Eawag, Swiss Federal Institute of Aquatic Sciences and Technology, Dubendorf, Switzerland.

Abdulmohsin, H., El-shafie, A., Hanna, W., Wan, M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. 541, 902–913. https://doi.org/10.1016/j.jhydrol.2016.07.048.

Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M., Kisi, O., 2019a. Daily streamflow prediction using optimally pruned extreme learning machine. J. Hydrol. 577. https://doi.org/10.1016/j.jhydrol.2019.123981.

Adnan, R.M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., Li, B., 2019b. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. J. Hydrol. 124371. https://doi.org/10.1016/j.jhydrol.2019.124371.

Ahmed, A.N., Othman, F.B., Afan, H.A., Elsha, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578, 1–18. https://doi.org/10.1016/j.jhydrol.2019. 124084.

Alizadeh, M.J., Kavianpour, M.R., 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. Mar. Pollut. Bull. 98, 171–178. https://doi.org/10.1016/j.marpolbul.2015.06.052.

Barzegar, R., Adamowski, J., Asghari, A., 2016. Application of wavelet artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. Stochastic Environ. Res. Risk Assess. 30, 1797–1819. https://doi.org/10.1007/s00477-016-1213-y.

Bauwe, A., Kahle, P., Lennartz, B., 2019. Evaluating the SWAT model to predict streamflow, nitrate loadings and crop yields in a small agricultural catchment. Adv. Geosci. 48, 1–9. https://doi.org/10.5194/adgeo-48-1-2019.

Chotpantarat, S., Boonkaewwan, S., 2018. Impacts of land-use changes on watershed discharge and water quality in a large intensive agricultural area in Thailand. Hydrol. Sci. J. 63, 1386–1407. https://doi.org/10.1080/02626667.2018.1506128.

Dagnew A, Philip T, Silvia WG, 2016 Modeling agricultural watersheds with the Soil and Water Assessment Tool (SWAT): Calibration and validation with a novel procedure for spatially explicit HRUs. Environmental Management 57, 894–911. doi:10.1007/s00267-015-0636-4.

Haghiabi AH, Nasrolahi AH, Parsaie A, 2018 Water quality prediction using machine learning methods. Water Quality Research Journal | 53.1, 3–13. doi:10.2166/wqrj.2018.025.

Herrig, I.M., Simone, I.B., Brennholt, N., Manz, W., 2015. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. Water Res. 85, 148–157. https://doi.org/10.1016/j.watres.2015.08.006.

Ho, J.Y., et al., 2019. Towards a time and cost-effective approach to water quality index class prediction. J. Environ. Manage. 575, 148–165. https://doi.org/10.1016/j.jhydrol.2019.05.016.

Hunter, J.M., Maier, H.R., Gibbs, M.S., Foale, E.R., Grosvenor, N.A., Harders, N.P., Kikuchi-miller, T.C., 2018. Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems. Hydrol. Earth Syst. Sci. 22, 2987–3006.

Isik, S., Kalin, L., Schoonover, J., Srivastava, P., Lockaby, B.G., 2013. Modeling effects of changing land use/cover on daily streamflow: an artificial neural network and curve number-based hybrid approach. J. Hydrol. 485, 103–112. https://doi.org/10.1016/j.jhydrol.2012.08.032.

Kalin, L., Isik, S., Lockaby, B.G., 2010. Predicting water quality in unmonitored watersheds using artificial neural networks. J. Environ. Qual. 39, 1429–1440.

Kavian, A., Mohammadi, M., Gholami, L., Rodrigo-comino, J., 2018. Assessment of the spatiotemporal effects of land use changes on runoff and nitrate loads in the Talar river. Water 10, 1–19. https://doi.org/10.3390/w10040445.

Keshtegar, B., Heddam, S., 2018. Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. Neural Comput. Appl. 30, 2995–3006. https://doi.org/10.1007/s00521-017-2917-8.

Khataar, M., Mosaddeghi, M.R., Amiri Chayjan, R., Mahboubi, A.A., 2018. Prediction of water quality effect on saturated hydraulic conductivity of soil by artificial neural networks. Paddy Water Environ, 16, 631–641.

Lee, S., Yeo, I.Y., Sadeghi, A.M., McCarty, G.W., Hively, W.D., Lang, M.W., Sharifi, A., 2018. Comparative analyses of hydrological responses of two adjacent watersheds to climate variability and change using the swat model. Hydrol. Earth Syst. Sci. 22, 689–708. https://doi.org/10.5194/hess-22-689-2018.

Libera, D.A., Sankarasubramanian, A., 2018. Multivariate bias corrections of mechanistic water quality model predictions. J. Hydrol. 564, 529–541. https://doi.org/10.1016/j.jhydrol.2018.07.043.

Malagó, A., Bouraoui, F., Vigiak, O., Grizzetti, B., Pastori, M., 2017. Modelling water and nutrient fluxes in the Danube River Basin with SWAT. Sci. Total Environ. 603–604, 196–218. https://doi.org/10.1016/j.scitotenv.2017.05. 242.

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. Am. Soc. Agric. Biol. Eng. 50 (3), 885–900.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models: Part I. A discussion of principles. J. Hydrol. 10, 282–290.

Neitsch SL, Arnold JG, Kiniry JR, Williams JR, 2011 Soil & Water Assessment Tool Theoretical Documentation Version 2009. Tech. rep., Texas Water Resources Institute Technical Report No. 406 Texas AM University System.

Niraula, R., Kalin, L., Srivastava, P., Anderson, C., 2013. Identifying critical source areas of nonpoint source pollution with SWAT and GWLF. Ecol. Model. 268, 123–133. https://doi.org/10.1016/j.ecolmodel.2013.08.007.

Noori, N., Kalin, L., 2016. Coupling SWAT and ANN models for enhanced daily streamflow prediction. J. Hydrol. 533, 141–151. https://doi.org/10.1016/j.jhydrol.2015.11.050.

Qi, M., Zhang, G., 2001. An investigation of model selection criteria for neural network time series forecasting. Eur. J. Oper. Res. 132, 666–680. https://doi.org/10.1016/S0377-2217(00)00171-5.

Rezaeianzadeh, M., Kalin, L., Anderson, C., 2015. Wetland water level prediction using ANN in conjunction with baseflow recession analysis. J. Hydrol. Eng. 22 (1). https://doi.org/10.1061/%28ASCE%29HE.1943-5584.0001276.

Rezaeianzadeh, M., Kalin, L., Hantush, M., 2018. An integrated approach for modeling wetland water level: application to a headwater wetland in coastal Alabama, USA. Water 10 (7), 879. https://doi.org/10.3390/w10070879.

Runkel BRL, Crawford CG, Cohn TA, 2004 LOAD ESTIMATOR (LOADEST): A Fortran program for estimating constituent loads in streams and rivers. U.S. Geological Survey, techniques and methods book 4, chapter a5 edn.

Salas J, Markus M, Tokar A, 2000 Streamflow forecasting based on artificial neural networks. In: Govindaraju R.S., Rao A.R. (eds) Artificial Neural Networks in Hydrology. Water Science and Technology Library 36, 23–51.

Sarkar, A., Pandey, P., 2015. River Water quality modelling using artificial neural network technique. Aquat. Proc. 4, 1070–1077. https://doi.org/10.1016/j.aqpro.2015.02.135.

Sefick, S., Kalin, L., Kosnicki, E., Schneid, B., Jarrell, M., Anderson, C., Paller, M., Feminella, W., 2015. Empirical estimation of stream discharge using channel geometry in low-gradient, sand-bed streams of the southeastern plains. J. Am. Water Resour. Assoc. 51, 1060–1071. https://doi.org/10.1111/jawr.12278.

Shi, Y., Xu, G., Wang, Y., Engel, B.A., Peng, H., Zhang, W., Cheng, M., Dai, M., 2017. Modelling hydrology and water quality processes in the Pengxi River basin of the Three Gorges Reservoir using the soil and water assessment tool. Agric. Water Manage. 182, 24–38. https://doi.org/10.1016/j.agwat.2016.12.007.

Šiljić Tomić, A., Antanasijević, D., Ristić, M., Perić-Grujić, A., Pocajt, V., 2018. Application of experimental design for the optimization of artificial neural network-based water quality model: a case study of dissolved oxygen prediction. Environ. Sci. Pollut. Res. 25, 9360–9370. https://doi.org/10.1007/s11356-018-1246-5.

Singh, H., Kalin, L., Morrison, A., Srivastava, P., Lockaby, B.G., Pan, S., 2015. Post-validation of SWAT model in a coastal watershed for predicting land use/cover change impacts. Hydrol. Res. 46 (6), 837–853. https://doi.org/10.2166/nh.2015.222.

Sirisha, P., Sravanti, K.N., Ramakrishna, V., 2008. Application of artificial neural networks for water quality prediction. Int. J. Syst. Technol. 1, 115–123.

Stidson, R.T., Gray, C.A., Mcphail, C.D., 2012. Development and use of modelling techniques for real-time bathing water quality predictions. Water Environ. J. 26, 7–18. https://doi.org/10.1111/j.1747-6593.2011.00258.x.

Tongal, H., Booij, M.J., 2018. Simulation and forecasting of stream flows using machine learning models coupled with base flow separation. J. Hydrol. 564, 266–282. https://doi.org/10.1016/j.jhydrol.2018.07.004.

Wallace, C.W., Flanagan, D.C., Engel, B.A., 2018. Evaluating the effects of watershed size on SWAT calibration. Water 10, 1–27. https://doi.org/10.3390/w10070898.

Wang, G., Jager, H.I., Baskaran, L.M., Baker, T.F., Brandt, C.C., 2016. Swat modeling of water quantity and quality in the Tennessee river basin: spatiotemporal calibration and validation. Hydrol. Earth Syst. Sci. Discuss. 2016, 1–33. https://doi.org/10.5194/hess-2016-34.

Wang, R., Kalin, L., 2011. Modeling effects of land use/cover changes under limited data. Eco-Hydrol. 4 (2), 265–276. https://doi.org/10.1002/eco.174.

Wu, Y., Shi, X., Li, C., Zhao, S., Pen, F., Green, T.R., 2017. Simulation of hydrology and nutrient transport in the Hetao irrigation district, inner Mongolia, China. Water 9, 1–15. https://doi.org/10.3390/w9030169.

Yan, T., Bai, J., Td, T., Arsenio, I.F., Liu, J., Shen, Z., 2019. Future climate change impacts on streamflow and nitrogen exports based on CMIP5 projection in the Miyun Reservoir Basin, China. Ecohydrol. Hydrobiol. 19, 266–278. https://doi.org/10.1016/j.ecohyd.2018.09.001.

Yaseen, Z.M., Ramal, M.M., 2018. Hybrid adaptive neuro-fuzzy models for water quality index estimation. Water Resour. Manage. 32, 2227–2245. https://doi.org/10.1007/s11269-018-1915-7.

Zhang, L., Zhang, G.X., Li, R.R., 2016. Water quality analysis and prediction using hybrid time series and neural network models. J. Agric. Sci. Technol. 18, 975–983.