

Comparison of short-term streamflow forecasting using stochastic time series, neural networks, process-based, and Bayesian models

Moges B. Wagena^a, Dustin Goering^b, Amy S. Collick^c, Emily Bock^a, Daniel R. Fuka^a, Anthony Buda^d, Zachary M. Easton^{a,*}

^a Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, USA

^b North Central River Forecast Center, NOAA National Weather Service, Minneapolis, MN, USA

^c Department of Agriculture, Food and Resource Sciences, University of Maryland Eastern Shore, Princess Anne, MD, USA

^d Pasture Systems and Watershed Management Research Unit, USDA-ARS, University Park, PA, USA

ARTICLE INFO

Keywords:

SWAT-VSA
ANNs
ARMA
Forecasting
Stochastic model
Process-based model
Bayesian model

ABSTRACT

Streamflow forecasts are essential for water resources management. Although there are many methods for forecasting streamflow, real-time forecasts remain challenging. This study evaluates streamflow forecasts using a process-based model (Soil and Water Assessment Tool-Variable Source Area model-SWAT-VSA), a stochastic model (Artificial Neural Network -ANN), an Auto-Regressive Moving-Average (ARMA) model, and a Bayesian ensemble model that utilizes the SWAT-VSA, ANN, and ARMA results. Streamflow is forecast from 1 to 8 d, forced with Quantitative Precipitation Forecasts from the US National Weather Service. Of the individual models, SWAT-VSA and the ANN provide better predictions of total streamflow (NSE 0.60–0.70) and peak flow, but underpredicted low flows. During the forecast period the ANN had the highest predictive power (NSE 0.44–0.64), however all three models underpredicted peak flow. The Bayesian ensemble forecast streamflow with the most skill for all forecast lead times (NSE 0.49–0.67) and provided a quantification of prediction uncertainty.

1. Introduction

Streamflow forecasting is essential for hydrologists, water resource planners, water system managers, emergency response providers, and policymakers to respond to ever-increasing water demand and greater variability. However, since streamflow is largely driven by stochastic processes (rainfall, temperature, seasonal variability) and complex, nonlinear watershed responses (Shao et al., 2009; Londhe and Charhate, 2010), it is challenging to precisely forecast streamflow, particularly as forecast lead times increase. Challenges to accurate streamflow forecasting as outlined by Pagano et al. (2014) include automating real-time data assimilation for model forcing, that rainfall-runoff models are simplifications of real-world processes, and that precipitation forecasts are inherently uncertain.

There are two broad rainfall-runoff modeling approaches used to forecast streamflow: physical/process-based models and empirical/statistical models (Masselot et al., 2016). Process-based models attempt to incorporate the relevant physical laws controlling watershed response and streamflow generation, generally requiring substantial effort to

build, parameterize, and calibrate, and often utilize an extensive amount of observed data. However, process-based models can provide the more reliable streamflow forecast compared to other forecasting techniques (Zealand et al., 1999). For instance, Demirel et al. (2009) compared the performance of a process-based model, the Soil and Water Assessment Tool (SWAT, Arnold et al., 1998) and an empirical artificial neural network (ANN) model in forecasting streamflow and found that, while ANNs were better at forecasting peak flow, SWAT performed better overall. Similarly, Hapuarachchi et al. (2011) found that process-based models provide more credible results in predicting flash flooding than empirical models. While process-based models can often forecast streamflow with adequate skill, there can also be tremendous uncertainty in the model predictions because of the uncertainty in input data, model structure, and the numerical techniques employed. Unfortunately, this uncertainty is often not included quantitatively in deterministic forecasts but is often incorporated into the final forecast product qualitatively using post processing tools, and forecaster experience.

Empirical models attempt to reproduce the relationships between

* Corresponding author.

E-mail address: zeaston@vt.edu (Z.M. Easton).

<https://doi.org/10.1016/j.envsoft.2020.104669>

Received 21 August 2019; Received in revised form 12 February 2020; Accepted 14 February 2020

Available online 17 February 2020

1364-8152/© 2020 Elsevier Ltd. All rights reserved.

inputs (e.g., precipitation) and outputs (streamflow) with no assumption of internal process understanding. Empirical models are often more easily developed and can provide reliable forecasts when the models are trained on robust, representative data. Several different types of empirical models have been used to forecast streamflow: including time series models such as the Auto Regressive Moving Average (ARMA) procedure (Babu and Reddy, 2014; Mohammad, 2015), ANNs (Khashei and Bijari, 2010; Kalteh, 2013), functional regressions (Massetot et al., 2016), wavelet support vectors (Kisi and Cimen, 2011), or decision trees (Bhattacharya and Solomatine, 2005).

Time series models, also known as auto-regressive models, can approximate complex non-linear hydrologic processes by using the relationship between observed values of streamflow and precipitation (Shao et al., 2009). Auto-regressive models assume that the present condition depends on the past conditions, often adjusted by an error function and a seasonality function (Yurekli et al., 2005). Time series models require a large amount of observed data to train and test the model (Yurekli et al., 2005; Khashei et al., 2009) and often perform poorly when applied outside of the data range for which they were developed (Guo et al., 2011). However, there have been many successful applications of auto-regressive models used in hydrologic forecasting. For instance, Toth et al. (2000) showed that using an ARMA model for real-time flood forecasting provided better predictive ability than simple rainfall-runoff models, emphasizing that empirical models can often outperform overly simplistic process based models. Collischonn et al. (2007) used an ARMA model to forecast medium-range reservoir inflow based on quantitative precipitation forecasts (QPF) and showed that prediction errors of an ARMA model could be reduced using improved rainfall forecasts.

Another type of empirical model commonly used to forecast streamflow is an ANN. ANNs are non-linear, data-driven, self-adaptive type models that can learn and generalize trends based on functional relationships in the data (Coulily et al., 2000; Zhang et al., 2001). The ANN method has been applied in many different sectors, including traffic control (Srinivasan et al., 2006), image processing (Kulkarni, 1994), stock forecasting (Guresen et al., 2011), handwriting analysis (Nasrabadi, 2007), and language recognition (Graves et al., 2013). In the water resources and environmental engineering fields, ANNs have been employed for rainfall forecasting (French et al., 1992; Luk et al., 2001), rainfall-runoff modeling (Shamseldin, 1997; Tokar and Johnson, 1999), reservoir operation (Jain et al., 1999; Rani and Moreira, 2010), and water quality forecasting (Palani et al., 2008; Singh et al., 2009). Although ANNs have been widely used in water resource fields there remain challenges, including the initial effort to build the model, computational power to run the model, and data required to train ANNs. Additionally, it can be difficult to identify important input variables, parameterize the number of ANN layers, and identify which transfer functions best describe the system (Zhang et al., 2001). One of the major shortcomings with ANNs is that they cannot work with major changes to the system, they must be re-trained to the new system state after major disturbances (Sudheer et al., 2003).

One other streamflow forecasting technique that is gaining popularity leverages the skill from multiple models to create an ensemble streamflow forecast. One significant advantage of the multiple model ensemble (MME) approach is the ability to evaluate model uncertainty (Weigel et al., 2008). Cloke and Pappenberger (2009) provide a review of studies employing MME approaches to exploit the diversity of skillful predictions from different models for river flood forecasting and enumerate the challenges related to the ensemble approach, such as computing power, resolution of input data, quantification and analysis of total uncertainty, communicating uncertainty, and probabilistic forecasting. One such MME approach is the Bayesian Generalized Multilevel Modeling (BGMM) method, which integrates multiple probability distributions by using prior knowledge about parameter distributions, allows predictor variables to be linear or nonlinear, and incorporates information from all predictors (models). The method also

offers a more complete quantification of uncertainty rather than simply attempting to improve forecasting skill (Wagena et al., 2019).

This study compares the performance of a process-based model, SWAT-VSA, with a modified ARMA time series model, an ANN model, and a BGMM application in a real-time forecast of streamflow. We applied each model in the US Department of Agriculture WE-38 experimental watershed in east-central Pennsylvania, US, and forecast streamflow from 1 to 8 days into the future using a hindcast procedure. Precipitation data from the US National Weather Service QPF was used to force each model for each forecast day. The BGMM was developed by ensembling the outputs from the SWAT-VSA, ARMA, and ANN models for each forecast day.

2. Materials and methods

2.1. Watershed description

The WE-38 experimental watershed is a sub-watershed of the Mahantango Creek watershed in east-central Pennsylvania, which drains to the Susquehanna River (Fig. 1). The watershed has an area of 7.3 km² and has been extensively studied as a United States Department of Agriculture Agricultural Research Service (USDA-ARS) experimental watershed since 1966 and contains a wealth of data to evaluate model performance. The watershed is characterized by complex terrain, with multiple runoff generating mechanisms, several crop types, and variable climate (snow dominated winters, humid temperate summers). Runoff generation and shallow subsurface flow through the vadose are controlled by highly fractured, shallow bedrock layers and runoff occurs primarily on variable saturated areas underlain by low permeability (fragipan) layers (Bryant et al., 2011). As part of the Appalachian Valley and Ridge Province, the watershed is underlain by fine siltstone, supporting soils prone to variable source area (VSA) runoff generation. The uplands feature well-drained soils with high infiltration rates while the lower landscape positions are more poorly drained soil with features that seasonally perch water and result in runoff generation by saturation excess processes (Lu et al., 2015). This combination of factors make WE38 an ideal location to test the suitability of models to forecast hydrologic responses across a highly variable watershed. The climate of WE-38 is temperate humid, with a mean temperature of 10.1°C, mean precipitation of 1080 mm yr⁻¹, and mean streamflow equal to 46% of precipitation (Buda et al., 2011; Lu et al., 2015). Elevation ranges from 220 to 510 m and the land use of the watershed consists of agriculture (44.5%), forest (33.8%), and pasture (3.5%) (Collick et al., 2015).

2.2. SWAT model description

The SWAT model is a process-based, semi-distributed watershed scale model developed to assess the impact of land management practices on water availability and water quality (Arnold et al., 1998). SWAT requires meteorological (precipitation, min and max temperature, solar radiation, wind speed, and humidity), soil, land cover, and land management data to simulate surface and subsurface hydrology and various chemical, nutrient, and sediment fluxes. SWAT-VSA re-conceptualizes SWAT to account for areas of the landscape subject to variable saturation dynamics. In SWAT-VSA the area of each hydrological Response Unit (HRU) is defined by the coincidence of land use and wetness index class, which is determined from a Topographic Index (TI), to differentiate areas of the landscape with respect to their moisture storage and saturation/runoff potential (Easton et al., 2008, 2011). SWAT-VSA was selected because it has been shown to provide better predictions of soil moisture, runoff generation, and nutrient export than the standard SWAT model in WE-38 and similar watersheds (Easton et al., 2008; Collick et al., 2015; Wagena et al., 2018). It is also capable of predicting these responses at a sub-field levels, which is important information for land managers and agricultural producers seeking to reduce the water quality impacts of human activities.

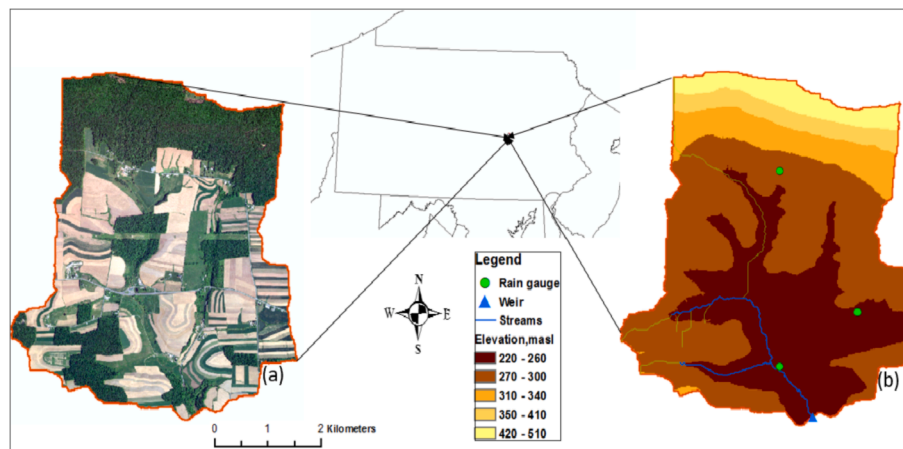


Figure 1. Location of WE-38 watershed in Pennsylvania, USA, and land cover ortho-imagery for 2015 (a) and digital elevation model (b) showing the location of the weir and precipitation gauges.

2.3. Watershed model initialization and input data

SWAT-VSA was initialized with a 10-m resolution digital elevation model resampled from 0.5 m LiDAR data obtained from Canaan Valley Institute (2007) using ArcSWAT 2012 and TopoSWAT (available from <https://dx.doi.org/10.6084/m9.figshare.1342823>) developed by Fuka et al. (2016). TopoSWAT automates the SWAT-VSA initialization process by creating the TI data and then overlaying the soil and TI data to develop the required database for model initialization. Soils data utilized by Topo SWAT are based on the Food and Agriculture Organization (FAO) soils database (FAO, 2007). This methodology downscales the FAO soils data, distributes the soil properties across TI classes, and has been shown to provide a more accurate representation of soil properties than SSURGO (Fuka et al. (2016)). The land use characterization of WE-38 was derived from previous studies (Gburek et al., 2002, 2006; Needelman et al., 2004; Veith et al., 2008; Buda et al., 2009, 2013). The model was initialized using measured precipitation, temperature (min and max), relative humidity, wind speed and solar radiation from 1987 to 2010 from land-based stations in WE-38 (Wagena et al., 2018).

2.4. Quantitative precipitation forecasts (QPFs)

The QPF is a forecast product developed by the US National Weather Service, Weather Prediction Center. The QPF is the expected amount of accumulated liquid precipitation, defined as the expected areal average on a 20×20 km grid in a given amount of time. The output is then post processed and downscaled to 5 km in final form to send to end users. The QPF data are generated in UTC, not in local time. The QPF data are adjusted for topographic effect. Additionally, forecasters adjust QPF data based on their experience to minimize the uncertainty of the forecast. It is also important to note that QPF data are provided in 6-hr accumulations so sub-6-hr intensity variability is not provided.

2.5. Model calibration and evaluation

The SWAT-VSA model was calibrated using SWAT-CUP (SWAT Calibration and Uncertainty Procedure) (Arnold et al., 2012) and SUFI2 (Sequential Uncertainty Fitting) as optimization algorithms with the objective function, which was set to the Nash Sutcliffe Efficiency coefficient (NSE). The SWAT-VSA model performance was evaluated based on three metrics: percent bias (PBIAS), root mean square error (RMSE), and the NSE. These three metrics were assessed using historical observations from two time periods: 1989 to 1998 for model calibration and 1999 to 2007 for model evaluation. The NSE is an indicator of the predictive power of the model and ranges from $-\infty$ to 1, where 1 is a perfect fit between modeled and observed data, and an NSE of

0 indicates that the observed mean provides a better estimate of the data than the model (Krause et al., 2005). The PBIAS is a statistical metric that provides an estimate of overprediction (PBIAS > 0) or underprediction (PBIAS < 0) of the model, while RMSE is a measure of the spread of observed values about the predicted values.

2.6. SWAT-VSA streamflow forecast

Quantitative precipitation forecasts for 1- to 8-day lead times were used as input to force the SWAT-VSA model in a forecast mode. Since QPF only contains precipitation data, the remainder of the required weather variables, such as relative humidity, solar radiation, wind speed, and temperature, were simulated using the SWAT weather generator (Sharpley and Williams, 1990). The daily values for temperature (maximum and minimum) and solar radiation were generated using a weekly stationary process described by Matalas (1967), relative humidity was generated using a triangular distribution from monthly averages (Sharpley and Williams, 1990), and wind speed was generated using a modified exponential equation (Sharpley and Williams, 1990). Two distinct sets of meteorological data were used in this study. The first was historical data used for model calibration and evaluation. The second, referred to as archived forecasts, was used to evaluate the model's capability to forecast without waiting months to collect future QPF forecast data. After model calibration and evaluation using the historic data the QPF data set prepared for the hindcast procedure was used to initialize unique SWAT-VSA model runs for each day in the hindcast period (January 1, 2017 to February 28, 2018; 444 days), recreating the live forecast procedure that the model was to perform. Prior to using the QPF forecast data to force the SWAT-VSA model, the data were converted from UTC to local time. During conversion of QPF data from UTC to local time, there was missing data of 7–8 h (depending on the forecast day) for the first forecast day. In order to ensure that this bias was minimized and did not propagate to the remaining forecasts, the missing data were replaced with observed precipitation data. In cases where no observed data were available, other sources of precipitation data such as satellite or gridded precipitation data sources (e.g., CFSR data) could be used to fill missing data. Each unique model run was forced with 12 years of observed meteorological data, with the last eight days of the run defined by the corresponding forecast day (i.e., 24 h, 48 h, 72 h, up to 192 h) from the QPF data. The remaining input was defined by the meteorological data prepared for calibration/evaluation, just as it would be in a live forecast.

The last eight days of each SWAT-VSA daily watershed discharge time series output, representing the hydrologic forecasts, was saved and separated into eight unique time series by forecast level. Thus, the streamflow hindcast datasets were generated as vectors of data points,

X_{ij} (one vector for each of the eight QPF lead times), where X_{ij} is the forecast of streamflow generated by model run i (1- y) for a specific lead time of the QPF forecast (1–8 days), j ; $i + j$ is equal to the day number within the hindcast period plus one (1- $y+1$). For example, the 24-hr archived forecasts vector is expressed as $[X_{1,1}, X_{2,1}, X_{3,1}, \dots, X_{y,1}]$, and the 72-hr archived forecasts vector as $[X_{1,1}, X_{1,2}, X_{1,3}, X_{2,3}, X_{3,3}, \dots, X_{y-2,3}]$. This procedure produced time series for all eight hindcast days that could be directly compared to measured volumetric flow at the watershed outlet.

2.7. Time series models including exogenous covariates

ARMA models typically consider the previous condition data (lag) and some error term to ‘forecast’ future conditions (Babu and Reddy, 2014), as is commonly used in reservoir forecasting. However, given the stochastic component of rainfall and its influence on streamflow generation, we modify the standard ARMA model form to include exogenous covariates, specifically precipitation. The general ARMA time series model is described by eq. (1):

$$y_t = \varnothing_1 y_{t-1} + \varnothing_2 y_{t-2} + \dots + \varnothing_p y_{t-p} + \beta' X_t + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1)$$

Fitting eq. (1) for the historical ARMA time series results in:

$$y_t = \varnothing_1 y_{t-1} + \varnothing_2 y_{t-2} + \varnothing_3 y_{t-3} + \beta' X_t + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} \quad (2)$$

The general ARMA forecast equation:

$$y_{t+m} = \varnothing_1 y_{t-1+m} + \varnothing_2 y_{t-2+m} + \varnothing_p y_{t-3+m} + \beta' X_t + a_{t+m} - \theta_1 a_{t-1+m} - \theta_2 a_{t-2+m} - \theta_q a_{t-3+m} - \theta_q a_{t-4+m} \quad (3)$$

The fitted ARMA forecast equation for lead time, one day ($m = 1$ day):

$$y_{t+1} = \varnothing_1 y_t + \varnothing_2 y_{t-1} + \varnothing_p y_{t-2} + \beta' X_t + a_{t+1} - \theta_1 a_t - \theta_2 a_{t-1} - \theta_q a_{t-2} - \theta_q a_{t-3} \quad (4)$$

Where y_t is the predicted streamflow (m^3/s) at time, t (d); y_{t-1} and y_{t-p} are previous data values for a given time $t-1$ and $t-p$; a_t is the error term at time, t ; a_{t-1} and a_{t-q} are previous error terms for time $t-1$ and $t-q$; \varnothing_1 , \varnothing_{t-p} , θ_1 , and θ_{t-q} are parameters to be fitted using observed data, p , and q are the order of the autoregressive and moving average components, respectively. X_t is a row vector containing the values of precipitation covariates at time t , β' is a column vector containing the coefficients related to those covariates, e.g., precipitation and m is the forecast lead time (d). The covariate is precipitation, which is included to enhance the forecast ability of ARMA time series model.

ARMA time series models have three components: the autoregressive (AR), the moving average (MA) components, and the exogenous covariate. To fit the regression, the observed streamflow dataset was split into two datasets, which corresponded to the SWAT-VSA calibration and evaluation periods. Prior to the fitting of the ARMA model, stationarity of the observed streamflow data was checked using the Augmented Dickey-Fuller method (Cheung and Lai, 1995), and the trend and seasonality of the streamflow data were checked using the decomposition method in R by checking for increasing or decreasing values in the time series for trend and by checking for patterns that repeat seasonally. All were found to be acceptable.

The ARMA model with exogenous covariates was first fit to the calibration dataset from 1989 to 1998 using eq. (1). During fitting of the ARMA model, the time series model orders (p , q) were derived from the autocorrelation functions (ACF) and partial autocorrelation functions

(PACF) by plotting ACF and PACF and checking whether there were significant autocorrelations with time lags in the flow data time series using R. The ACF is the linear dependence of a variable with itself at two points in time and is useful in determining whether a time series is stationary. The PACF is the autocorrelation between variables after removing linear dependence (Berryman & Turchin, 2001). After fitting the model using the calibration dataset, the best model was selected by checking the ACF and PACF residuals of the fitted model and determining whether the residuals of fitted model were normally distributed using Shapiro-Wilk test and quantile-quantile (Q-Q) plot. The ARMA model performance was then evaluated using separate streamflow and precipitation datasets from 1999 to 2007 period. Finally, the fitted ARMA time series model with the QPF forecast data as covariates (eq. (3)) was used to forecast streamflow for 1- to 8-day lead times for the 01/01/2017 through 02/28/2018 forecast period. Eight time series of forecast streamflow data, one for each forecast day, were extracted and compared with the observed streamflow data.

2.8. Artificial neural networks (ANNs)

The ANN model developed in this study was a deep feed forward neural network composed of an input layer, hidden layers, and an output layer (Schmidhuber, 2015). Layers are made up of nodes called neurons, which produce real-valued activations from a number of weighed connections to neurons from the previous layer (Schmidhuber, 2015). The first layer of an ANN model is made up of input neurons, which hold the value of the independent (predictor) variable, one variable per node (Fig. 2). This process is repeated through all layers and nodes, and a final computation is done to calculate the values for the dependent (prediction) variables, constituting the output layer (Schmidhuber, 2015). Training the ANN model involved incrementally adjusting the weights (eq. (5)) assigned to each connection (neuron) between nodes to optimize a loss function. The backpropagation algorithm implemented in this study (eq. (6)) is a comparatively fast and efficient training algorithm and currently very common among modern ANN code libraries (LeCun et al., 2015).

$$F_j = \sum w_{ji} x_i \quad (5)$$

$$\text{Output} = f(F_j) \quad (6)$$

Where F_j is the weighted sum of inputs, w are the weights, x are the inputs, and f is the activation function, j and i are indices for weights that links inputs, hidden and output layers. The most critical inputs to train the ANN were found to be the lag streamflow and observed

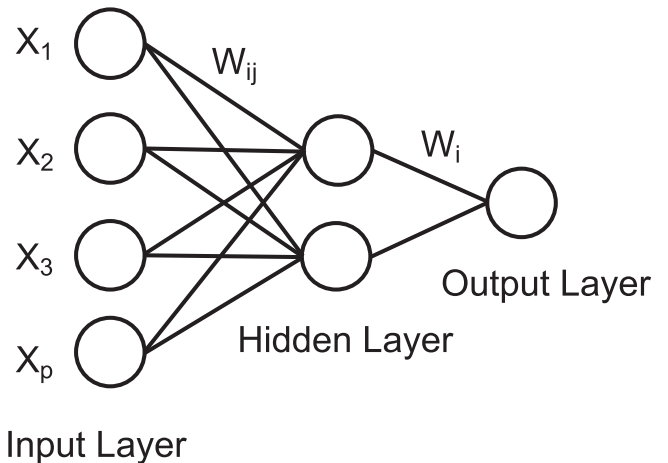


Fig. 2. Typical structure of an artificial neural network, X (inputs), and W (weights).

precipitation to train the networks. The input variables were selected by a trial and error method from a combination of different input variables (e.g., flow and precipitation) until weights converge by increasing the number of iterations and changing the transfer functions, after which performance was evaluated by comparing the fitted ANN output to observed streamflow. The number of hidden layers (four) was iteratively determined based on the performance of fitted ANN output. To implement the forecast in the ANN, a similar procedure to SWAT-VSA forecasting was developed, with the additional variable of one-day lag streamflow and precipitation from QPF used as input to the fitted network. The forecasted streamflow for the eight-day forecast time series was compared with the observed data to evaluate the performance of the ANNs.

2.9. Bayesian generalized (non-) linear multilevel models (BGMM)

The BGMM procedure, a statistical post-processing method, was used to ensemble the streamflow forecasts of the three models for each forecast period. The BGMM was fitted using the Stan Package in R (Buerkner, 2016; Carpenter et al., 2016; Wagena et al., 2019). Model predictions of streamflow for each forecast day from SWAT-VSA, ANN, and ARMA were used as BGMM input. The BGMM predicts a response, y , using a combination of predictors (three model outputs) transformed by an inverse link function, assuming a certain distribution, D using eqs. (7) and (8).

$$y_i \sim D(f(\eta_i), \theta) \tag{7}$$

$$\eta = \beta X + Z \tag{8}$$

where y_i is the response at time i , obtained through the linear combination of η of predictors transformed by inverse function, f , and assuming a certain distribution, D . β and u are fitted coefficients at the population level and group level, respectively, and X and Z are the corresponding design matrices. The model parameters β , u , and θ were estimated using the Hamiltonian Monte Carlo-NUTS algorithm (Buerkner, 2016) by defining the prior distributions and X (or X , Y , and Z , respectively). The weights placed on each of the model predictions were obtained by summing the absolute value of the fitted coefficient values, β , and then dividing each coefficient by the sum of the absolute value of the coefficients obtained during the training period, eq (9):

$$Weights(i) = \frac{|\beta_i|}{\sum_i |\beta_i|} \tag{9}$$

where the sum of all weights across i is equal to one, and coefficient, β are fitted absolute coefficient values of each predictor.

Finally, to determine if the models were able to maintain a component mass balance, we compared estimates of baseflow and runoff from each model against measured baseflow and runoff in the watershed. To separate baseflow and runoff for the ARMA, ANN, BGMM, and measured data we employed a simple signal filter (Luo et al., 2012), and for SWAT-VSA we used the model predictions of baseflow and runoff directly from the model.

Table 1

Percent bias (PBIAS), Root mean square error (RMSE) in m^3/s , and Nash Sutcliffe Efficiency coefficient (NSE) values for daily model calibration, and evaluation periods for WE38 watershed.

Response	ARMA		SWAT-VSA		ANN	
	calibration	evaluation	calibration	evaluation	calibration	evaluation
PBIAS	0.90	7.70	5.60	6.90	0.40	3.30
RMSE	0.21	0.18	0.20	0.11	0.17	0.12
NSE	0.40	0.35	0.60	0.70	0.60	0.65

3. Results

3.1. Historical model performance assessment

According to performance criteria recommended by Moriasi et al. (2007, 2015), a model simulation can be judged as satisfactory if $NSE > 0.5$ and $PBIAS \pm 25\%$ for streamflow. Based on these criteria, the results indicate that all three models predicted the observed flow well or moderately well during the calibration and evaluation periods (Table 1), with a low PBIAS and good NSE and RMSE performance metrics. All models slightly over-predicted the overall streamflow mass balance during calibration and evaluation periods as indicated by the PBIAS metric (Table 1), although peak flows were slightly underestimated (Fig. 3).

Fig. 3 shows observed and predicted daily streamflow for each model during the calibration and evaluation periods. All models simulated the daily low flows moderately well during both periods (Table 1, Fig. 3). Overall, SWAT-VSA and the ANN model provided the most accurate prediction of daily streamflow, with NSE of 0.60 or greater during both the training and testing periods (Table 1). All models tended to under-predict peak flows during both the training and testing periods, although less so for the ANN (Fig. 3). For baseflow, the ARMA model performed best, followed by the SWAT-VSA and ANN models (Fig. 3).

3.2. Forecast performance evaluation

The short-term streamflow forecasts for the three models were compared to measured data to assess the forecast ability of the models. The forecast results show all models were generally able to forecast streamflow well or moderately well during most lead times (Table 2, Figs. 4–7). As would be expected, all models exhibited better predictive performance during shorter lead times (24–72 h) compared to longer times (Table 2). All individual models generally underpredicted streamflow during all lead times and exhibited increasing prediction bias at greater lead times. The BGMM overpredicted during all lead time periods except during 72–120hr forecast periods when it underpredicted

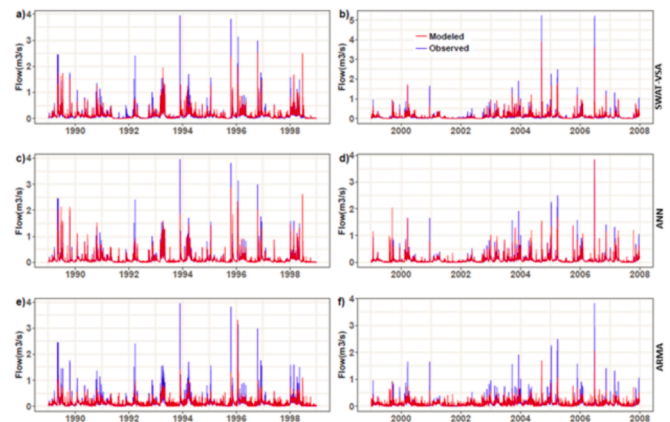


Fig. 3. Daily streamflow time series plot during the training and testing periods for SWAT-VSA (a, b), the ANN (c, d), and the ARMA (e, f) models.

Table 2

Nash Sutcliffe Efficiency coefficient (NSE), Percent Bias (PBIAS), and Root Mean Square Error (RMSE) in m³/s values for hourly forecast periods for WE38 watershed.

Time (hr)	SWAT-VSA			ARMA			ANN			BGMM		
	NSE	Pbias	RMSE	NSE	Pbias	RMSE	NSE	Pbias	RMSE	NSE	Pbias	RMSE
24	0.42	-13.20	0.11	0.49	-3.80	0.11	0.64	-0.90	0.09	0.67	2.60	0.09
48	0.44	-8.20	0.11	0.49	-4.10	0.11	0.60	-3.80	0.09	0.63	4.50	0.09
72	0.39	-1.90	0.12	0.48	-3.80	0.11	0.54	2.40	0.10	0.59	-1.40	0.10
96	0.27	-8.20	0.13	0.46	-4.60	0.11	0.55	-4.80	0.10	0.56	-0.20	0.10
120	0.21	-13.90	0.13	0.45	-4.40	0.11	0.44	-4.10	0.11	0.49	-1.90	0.11
144	0.21	-20.70	0.13	0.49	-4.80	0.11	0.58	-8.90	0.10	0.60	3.00	0.09
168	0.28	-20.40	0.13	0.50	-4.90	0.10	0.60	-8.80	0.09	0.62	4.10	0.09
192	0.25	-29.00	0.13	0.52	-5.50	0.10	0.56	-9.10	0.10	0.59	8.10	0.09

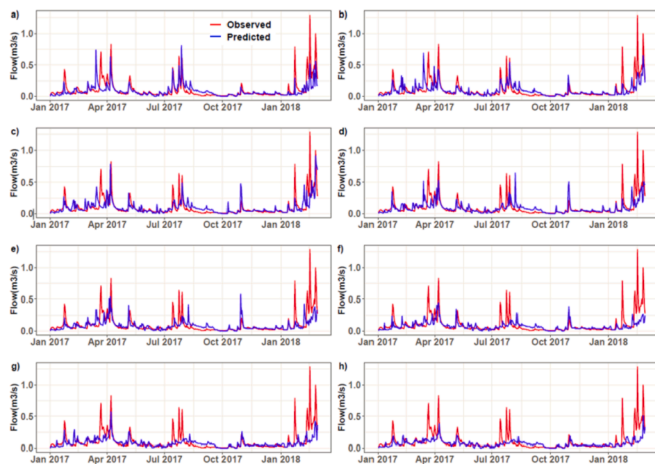


Fig. 4. Performance of the SWAT-VSA model for 24hr (a), 48hr (b), 72hr(c), 96hr (d), 120hr (e), 144hr (f), 168hr (g), and 192hr (h) streamflow forecasts.

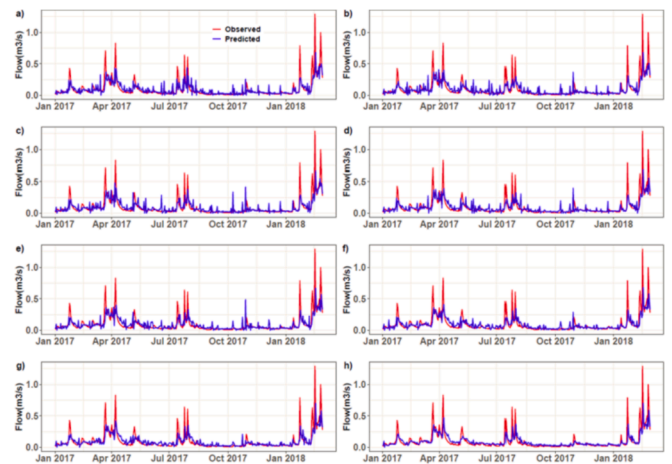


Fig. 6. Performance of the ARMA time series for 24hr (a), 48hr (b), 72hr(c), 96hr (d), 120hr (e), 144hr (f), 168hr (g), and 192hr (h) streamflow forecasts.

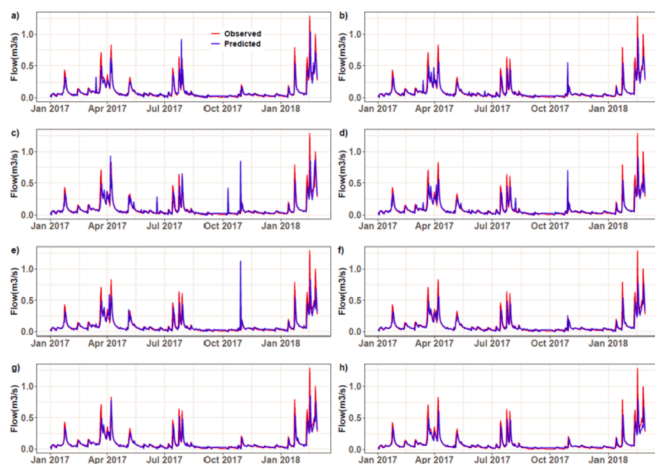


Fig. 5. Performance of the ANN model for 24hr (a), 48hr (b), 72hr(c), 96hr (d), 120hr (e), 144hr (f), 168hr (g), and 192hr (h) streamflow forecasts.

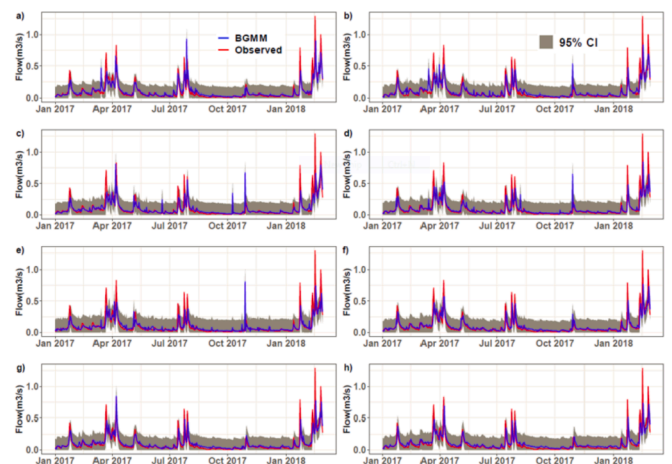


Fig. 7. Performance of the BGMM for 24hr (a), 48hr (b), 72hr (c), 96hr (d), 120hr (e), 144hr (f), 168hr (g), and 192hr (h) streamflow forecasts.

streamflow. The BGMM consistently exhibited the greatest NSE across all forecast lead times although the ANN produced similar forecast results.

Fig. 4 compares the streamflow forecast by SWAT-VSA to the observed streamflow for each forecast period. SWAT-VSA closely captured the trend of daily forecasted streamflow but tended to miss some peak flow events, particularly during the 120–192hr lead-times (Fig. 4). SWAT-VSA performed better in capturing peaks during the 24–96hr lead times, although there was still a tendency to underpredict high flows. While high flows were slightly underpredicted, the timing of events was well captured (Fig. 4). However, as indicated by Fig. 8,

SWAT-VSA correctly partitioned total streamflow between baseflow and runoff fractions, both during the calibration and evaluation and, for the most part, for the forecast period. The notable exception was for the 120–192hr forecasts, where SWAT-VSA underpredicted runoff (also corroborated in Fig. 4 and Table 2). The majority of this underprediction is due to the bias in QPF forecast precipitation falling well below the observed precipitation during the 120–192 h periods (Fig. 9).

The ANN model forecasted streamflow well across all lead times (Fig. 5). The ANN performed well in predicting both base and peak streamflow and the trend of the observed streamflow across all lead times, although there was a slight tendency to under-predict peak flow



Fig. 8. Average daily baseflow and runoff (in mm) for WE38 watershed during calibration (a), evaluation (b), forecasting: 24hr (c), 48hr (d), 72hr (e), 96hr (f), 120hr (g), 144hr (h), 168hr (i), and 192hr (j). Note that the BGMM was not run for the training and testing periods.

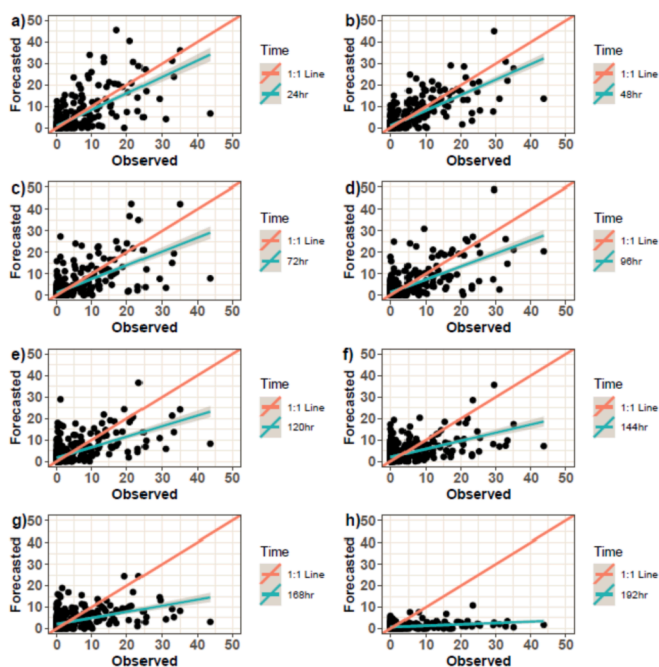


Fig. 9. Scatter plots of measured vs QPF forecast precipitation for the 24–192 h forecasts in WE38.

particularly during spring snowmelt (March–May in Fig. 5). One of the strengths of the ANN model is its significant performance skill in capturing baseflow during all lead times (Figs. 5 and 8). However, baseflow separation of the ANN predicted streamflow showed it underpredicted runoff during forecast periods except for the 72hr forecast period (Fig. 8). Interestingly, the systematic underprediction of QPF precipitation during the 120–192hr forecasts (Fig. 9) did not impact the ANN predictions nearly as dramatically as the process based SWAT-VSA model.

Fig. 6 shows the time series performance of the ARMA model for the forecast period. The ARMA model tended to underpredict peak flows and overpredict baseflows. Indeed, the mass balance in Fig. 8 indicates that the model overpredicted baseflow and underpredicted runoff. There is no apparent forecast lead time or time of year during which the ARMA

model performed better or worse.

The multi-model BGMM forecasted streamflow well during all lead times compared to the individual models (although the ANN produced similar results). Fig. 7 shows the time series performance of BGMM during the forecast period. The BGMM captured both baseflow and peak flow with skills similar to the ANN (Table 2), and the mass balance in Fig. 8 corroborates this. Also shown in Fig. 7 are the 95% credible intervals for the BGMM model predictions, which indicate the probability that the observed streamflow lies within the credible interval 95% of the time. As is clear from Fig. 7 and Table 3, the majority of the observed streamflow fell within the 95% credible interval of the BGMM, with only 3.1–4.2% of the observed data (depending on forecast hr) not contained by the interval. Moreover, it was the larger peaks that tended to fall outside of this interval.

4. Discussion

The performance variability among the models in simulating daily streamflow, baseflow, and runoff is of interest to explore. Process-based models with representative model forcing data, adequate physical process representation, and good structure that are well calibrated generally predict streamflow well compared to empirical models due to the process-based model’s representation of internal processes (Wang et al., 2011). However, well-defined and trained empirical models like ANNs have the ability to forecast streamflow with comparable or superior skill to process-based models, though the performance of ANNs is highly dependent on the number of layers and the transfer functions used to train the model (Maier et al., 2010). The availability of data is also crucial to generalize the performance of ANNs (Dawson and Wilby,

Table 3

Percent of days the observed streamflow falls outside of the 95% credible interval from the BGMM for each forecast hr.

Time	% of time outside of 95% CI
24	4.0
48	3.8
72	3.8
96	4.2
120	3.5
144	4.0
168	3.1
192	3.5

1998); ANNs are nonlinear in nature and with important variables and associated data, can be trained to learn and generalize complex problems. In contrast, fully empirical time series models, like the ARMA, are linear in nature and therefore often fail to capture complex, non-linear processes.

Results of this study run counter to some of the literature that shows a process-based model like SWAT generally outperforming empirical models in forecasting streamflow (Demirel et al., 2009). However, the inherent uncertainty stemming from model parameters, structure, mathematical process representations, inputs, and initial conditions can also cause process-based models to perform poorly compared to less parametrized empirical models during forecast periods (Block et al., 2009). Indeed, in this application the ANN model tended to produce the most accurate predictions (Table 2, Fig. 5). The ANN for WE38 watershed was trained with adequate data and a sufficient number of layers to simulate baseflow and runoff well during the calibration and evaluation periods, with predictive power similar to or better than SWAT-VSA (Table 1, Fig. 3c & d). The ARMA model also performed well in predicting the trend in streamflow, baseflow, and runoff when exogenous covariates were used (Table 1, Fig. 3e & d). However, while the ARMA model forecast streamflow moderately well, it more severely over-predicted baseflow and under-predicted runoff during all periods compared to the other two models (Fig. 8). SWAT-VSA simulated daily streamflow very well during the calibration and evaluation periods (Table 1, Fig. 3a & b) but, like most other process-based models, missed capturing some of the peaks due to the stochastic nature of streamflow. In contrast, the SWAT-VSA performance declined substantially during the forecast period from its performance during calibration and evaluation; its performance was lower than the other two individual models (Table 1, Table 2).

It is important to emphasize that the SWAT-VSA model was developed by considering the relevant processes that can be predicted at the landscape/field level using hydrologic response units. Modeling multiple, complex processes inherently introduces more sources of uncertainty than modeling a single output variable, as in the ANN and ARMA models. In addition, the SWAT model is known to simulate landscape processes quite well but not always subsurface processes, e.g., baseflow (Luo et al., 2012), which may account for its relatively high bias during calibration and evaluation (Table 1) and during the forecast period (Table 2), particularly for longer QPF lead times. Despite these shortcomings, the utility of SWAT-VSA goes well beyond streamflow simulation and forecasting. One of the benefits of SWAT-VSA over empirical models is its ability to make spatial predictions within the watershed for processes such as soil moisture, runoff generation, or nutrient export. For watershed management concerns beyond water supply and flooding, such as selection and targeting of best management practices, SWAT-VSA has high potential to enhance decision-making (Wagena and Easton, 2018). However, the process modeling that provides useful information for decision-making, particularly in agroecosystem management (e.g., plant growth, evapotranspiration, and soil moisture content), also affects the water balance, and can cause the uncertainty in each of these processes to cascade through the model and impact streamflow estimation. As a result, SWAT-VSA is likely to be more sensitive to bias in the QPF data (e.g., Fig. 9). Fig. 9 compares the observed precipitation to the QPF predictions and demonstrates that the QPFs had significant bias, systematically under predicting precipitation for all lead times, most acutely beyond 120hr.

Interestingly the bias in the QPF (Fig. 9) forecast did not impact the ANN or ARMA predictions in any systematic manner, they both provided approximately equivalent forecast predictions across all lead times, as evidenced in Table 2. This at first inspection seems counter intuitive, however, because both the ANN and ARMA models include a lag (observed) streamflow term, the models both are able to essentially bias correct the QPF data while running the forecast.

To overcome the shortcomings of any individual model, the BGMM leveraged the skill of three individual models and resulted in good

performance of streamflow forecasting during all forecast periods. The method considers the strengths and weaknesses of each model's performance in forecasting streamflow based on evaluating individual model performance and providing the 95% credible interval. Consequently, the ensemble of the three models' outputs using the BGMM approach generally outperformed the individual models in forecasting streamflow although the ANN results were similar (Table 2 & Fig. 7). In addition to the improved forecast skill exhibited by the BGMM approach, the uncertainty associated with the streamflow forecast is also provided, which can assist in risk-based decision making (Block et al., 2009). Since there are no perfect forecast models (Block et al., 2009), the BGMM approach could improve forecast skill with respect to runoff, baseflow, and associated cumulative uncertainty. Indeed, BGMM forecasted baseflow and runoff compared favorably to the individual models during all forecast periods by leveraging each of the model's forecasting abilities (Fig. 8).

5. Conclusions

Using forecasted streamflow from both process-based and empirical models helps water resources planning and management efforts. Streamflow forecasting, specifically, has an essential role in water resources and environmental engineering. For example, hydrologic forecasting is used by farmers to assess soil moisture condition and enable them to identify areas prone to runoff and erosion (Sommerlot et al., 2016), reducing flood damage by providing a warning period where actions can be taken to mitigate the impact, and by reservoir operators to ensure adequate water supplies (Shiri and Kisi, 2010). Overall, our results show that there is no single superior model that can be used to forecast streamflow, and each model has its own strengths and weaknesses. The ANN was able to learn stochastic, complex, nonlinear problems, although they still missed capturing some peak flows. While the process-based model did not always perform better than the empirical models, it does have the ability to provide spatial responses in the watershed and predict water quality both at the watershed outlet and at given locations along the stream network. ARMA models require relatively little data but are limited in their ability to simulate complex processes.

Declaration of competing interest

The authors have no conflicts of interest.

Acknowledgments

We would like to acknowledge high-performance computing support from Yellowstone (<http://n2t.net/ark:/85065/d7wd3xhc>) provided by NCAR's Computational and Information Systems Laboratory, support from the National Science Foundation under award numbers 1360415 and 1343802, and funding support from the USDA under project number 2012-67019-19434. All data and methods used in this manuscript are available upon request.

References

- Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R., Van Griensven, A., Van Liew, M.W., 2012. SWAT: model use, calibration, and validation. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 55 (4), 1491–1508.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment Part I: model development. *J. Am. Water Resour. Assoc.* 34 (1), 73–89. <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>.
- Babu, C.N., Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Appl. Soft Comput.* 23, 27–38. <https://doi.org/10.1016/j.asoc.2014.05.028>.
- Berryman, A., Turchin, P., 2001. Identifying the density-dependent structure underlying ecological time series. *Oikos* 92 (2), 265–270.

- Bhattacharya, B., Solomatine, D.P., 2005. Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing* 63, 381–396. <https://doi.org/10.1016/j.neucom.2004.04.016>.
- Block, P.J., Souza Filho, F.A., Sun, L., Kwon, H.-H., 2009. A streamflow forecasting framework using multiple climate and hydrological Models. *JAWRA J. Am. Water Resour. Assoc.* 45 (4), 828–843. <https://doi.org/10.1111/j.1752-1688.2009.00327>.
- Bryant, R.B., Veith, T.L., Feyereisen, G.W., Buda, A.R., Church, C.D., Folmar, G.J., Schmidt, J.P., Dell, C.J., Kleinman, P.J.A., 2011. U.S. Department of agriculture agricultural research Service Mahantango Creek watershed, Pennsylvania, United States: physiography and history. *Water Resour. Res.* 47 (8) <https://doi.org/10.1029/2010WR010056> n/a-n/a.
- Buda, A.R., Kleinman, P.J., Feyereisen, G.W., Miller, D.A., Knight, P.G., Drohan, P.J., Bryant, R.B., 2013. Forecasting runoff from Pennsylvania landscapes. *J. Soil Water Conserv.* 68 (3), 185–198.
- Buda, A.R., Kleinman, P.J., Srinivasan, M., Bryant, R.B., Feyereisen, G.W., 2009. Factors influencing surface runoff generation from two agricultural hillslopes in central Pennsylvania. *Hydrol. Process.* 23 (9), 1295–1312.
- Buda, A.R., Veith, T.L., Folmar, G.J., Feyereisen, G.W., Bryant, R.B., Church, C.D., Schmidt, J.P., Dell, C.J., Kleinman, P.J., 2011. U.S. Department of agriculture agricultural research Service Mahantango Creek watershed, Pennsylvania, United States: long-term precipitation database. *Water Resour. Res.* 47 (8) <https://doi.org/10.1029/2010WR010058> n/a-n/a.
- Buerkner, P.-C., 2016. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Software* 80 (1), 1–28.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., Riddell, A., 2016. Stan: a probabilistic programming language. *J. Stat. Software* 20, 1–37.
- Cheung, Y.-W., Lai, K.S., 1995. Lag order and critical values of the augmented Dickey–Fuller test. *J. Bus. Econ. Stat.* 13 (3), 277–280.
- Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. *J. Hydrol.* 375 (3), 613–626. <https://doi.org/10.1016/j.jhydrol.2009.06.005>.
- Collick, A.S., Fuka, D.R., Kleinman, P.J., Buda, A.R., Weld, J.L., White, M.J., Veith, T.L., Bryant, R.B., Bolster, C.H., Easton, Z.M., 2015. Predicting phosphorus dynamics in complex terrains using a variable source area hydrology model. *Hydrol. Process.* 29 (4), 588–601.
- Collischonn, W., Morelli Tucci, C.E., Clarke, R.T., Chou, S.C., Guilhon, L.G., Cataldi, M., Allasia, D., 2007. Medium-range reservoir inflow predictions based on quantitative precipitation forecasts. *J. Hydrol.* 344 (1), 112–122. <https://doi.org/10.1016/j.jhydrol.2007.06.025>.
- Coulibaly, P., Anctil, F., Bobée, B., 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* 230 (3), 244–257. [https://doi.org/10.1016/S0022-1694\(00\)00214-6](https://doi.org/10.1016/S0022-1694(00)00214-6).
- Dawson, C.W., Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.* 43 (1), 47–66.
- Demirel, M.C., Venancio, A., Kahya, E., 2009. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. *Adv. Eng. Software* 40 (7), 467–473. <https://doi.org/10.1016/j.advengsoft.2008.08.002>.
- Easton, Z.M., Fuka, D.R., Walter, M.T., Cowan, D.M., Schneiderman, E.M., Steenhuis, T.S., 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J. Hydrol.* 348 (3), 279–291. <https://doi.org/10.1016/j.jhydrol.2007.10.008>.
- Easton, Z.M., Walter, M.T., Fuka, D.R., White, E.D., Steenhuis, T.S., 2011. A simple concept for calibrating runoff thresholds in quasi-distributed variable source area watershed models. *Hydrol. Process.* 25 (20), 3131–3143.
- FAO, 2007. State of the World's Forests 2007. Food & Agriculture Org.
- French, M.N., Krajewski, W.F., Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network. *J. Hydrol.* 137 (1–4), 1–31.
- Fuka, D.R., Collick, A.S., Kleinman, P.J., Auerbach, D., Harmel, D., Easton, Z.M., 2016. Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model. *Hydrol. Process.* <https://doi.org/10.1002/hyp.10899>.
- Gburek, W., Drungil, C., Srinivasan, M., Needelman, B., Woodward, D., 2002. Variable-source-area controls on phosphorus transport: bridging the gap between research and design. *J. Soil Water Conserv.* 57 (6), 534–543.
- Gburek, W., Needelman, B., Srinivasan, M., 2006. Fragipan controls on runoff generation: hydrogeological implications at landscape and watershed scales. *Geoderma* 131 (3), 330–344.
- Graves, A., Mohamed, A.-r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. Paper presented at the 2013.
- Guo, J., Zhou, J., Qin, H., Zou, Q., Li, Q., 2011. Monthly streamflow forecasting based on improved support vector machine model. *Expert Syst. Appl.* 38 (10), 13073–13081. <https://doi.org/10.1016/j.eswa.2011.04.114>.
- Guresen, E., Kayakutlu, G., Daim, T.U., 2011. Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.* 38 (8), 10389–10397.
- Hapuarachchi, H.A.P., Wang, Q.J., Pagano, T.C., 2011. A review of advances in flash flood forecasting. *Hydrol. Process.* 25 (18), 2771–2784. <https://doi.org/10.1002/hyp.8040>.
- Jain, S., Das, A., Srivastava, D., 1999. Application of ANN for reservoir inflow prediction and operation. *J. Water Resour. Plann. Manag.* 125 (5), 263–271.
- Kalteh, A.M., 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput. Geosci.* 54, 1–8. <https://doi.org/10.1016/j.cageo.2012.11.015>.
- Khashei, M., Bijari, M., 2010. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Syst. Appl.* 37 (1), 479–489. <https://doi.org/10.1016/j.eswa.2009.05.044>.
- Khashei, M., Bijari, M., Raissi Ardali, G.A., 2009. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing* 72 (4), 956–967. <https://doi.org/10.1016/j.neucom.2008.04.017>.
- Kisi, O., Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J. Hydrol.* 399 (1–2), 132–140.
- Krause, P., Boyle, D., Båse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97.
- Kulkarni, A.D., 1994. Artificial Neural Networks for Image Understanding. *VNR Computer Library: Van Nostrand Reinhold*, New York.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436.
- Londhe, S., Charhate, S., 2010. Comparison of data-driven modelling techniques for river flow forecasting. *Hydrol. Sci. J.* 55 (7), 1163–1174. <https://doi.org/10.1080/02626667.2010.512867>.
- Lu, H., Bryant, R.B., Buda, A.R., Collick, A.S., Folmar, G.J., Kleinman, P.J.A., 2015. Long-term trends in climate and hydrology in an agricultural, headwater watershed of central Pennsylvania, USA. *J. Hydrol.: Reg. Stud.* 4 (Part B), 713–731. <https://doi.org/10.1016/j.ejrh.2015.10.004>.
- Luk, K.C., Ball, J.E., Sharma, A., 2001. An application of artificial neural networks for rainfall forecasting. *Math. Comput. Model.: Int. J.* 33 (6–7), 683–693.
- Luo, Y., Arnold, J., Allen, P., Chen, X., 2012. Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. *Hydrol. Earth Syst. Sci.* 16 (4), 1259–1267.
- Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ. Model. Software* 25 (8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>.
- Masselot, P., Dabo-Niang, S., Chebana, F., Ouarda, T.B.M.J., 2016. Streamflow forecasting using functional regression. *J. Hydrol.* 538, 754–766. <https://doi.org/10.1016/j.jhydrol.2016.04.048>.
- Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. *Water Resour. Res.* 3 (4), 937–945.
- Mohammad, V., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorol. Appl.* 22 (3), 592–598. <https://doi.org/10.1002/met.1491>.
- Moriassi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 50 (3). <https://doi.org/10.13031/2013.23153>.
- Moriassi, D.N., Zeckoski, R.W., Arnold, J.G., Baffaut, C., Malone, R.W., Daggupati, P., Guzman, J.A., Saraswat, D., Yuan, Y., Wilson, B.N., 2015. Hydrologic and water quality models: Key calibration and validation topics. *Transactions of the ASABE* 58 (6), 1609–1618.
- Nasrabadi, N.M., 2007. Pattern recognition and machine learning. *J. Electron. Imag.* 16 (4), 049901.
- Needelman, B.A., Gburek, W.J., Petersen, G.W., Sharpley, A.N., Kleinman, P.J., 2004. Surface runoff along two agricultural hillslopes with contrasting soils. *Soil Sci. Soc. Am. J.* 68 (3), 914–923.
- Pagano, T.C., Wood, A.W., Ramos, M.-H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Soroshian, S., Verkade, J.S., 2014. Challenges of operational river forecasting. *J. Hydrometeorol.* 15 (4), 1692–1707. <https://doi.org/10.1175/JHM-D-13-0188.1>.
- Palani, S., Liong, S.-Y., Tkalich, P., 2008. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* 56 (9), 1586–1597.
- Rani, D., Moreira, M.M., 2010. Simulation-optimization modeling: a survey and potential application in reservoir systems operation. *Water Resour. Manag.* 24 (6), 1107–1138.
- Schmidhuber, J., 2015. Deep learning in neural networks: an overview. *Neural Network.* 61, 85–117.
- Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modelling. *J. Hydrol.* 199 (3–4), 272–294.
- Shao, Q., Wong, H., Li, M., Ip, W.-C., 2009. Streamflow forecasting using functional-coefficient time series model with periodic variation. *J. Hydrol.* 368 (1), 88–95. <https://doi.org/10.1016/j.jhydrol.2009.01.029>.
- Sharpley, A., Williams, J., 1990. EPIC Erosion Productivity Impact Calculator: 2. In: *User Manual*, vol. 1768. US Department of Agriculture Technical Bulletin, p. 127.
- Shiri, J., Kisi, O., 2010. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *J. Hydrol.* 394 (3), 486–493. <https://doi.org/10.1016/j.jhydrol.2010.10.008>.
- Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality—a case study. *Ecol. Model.* 220 (6), 888–895.
- Sommerlot, A.R., Wagena, M.B., Fuka, D.R., Easton, Z.M., 2016. Coupling the short-term global forecast system weather data with a variable source area hydrologic model. *Environ. Model. Software* 86, 68–80. <https://doi.org/10.1016/j.envsoft.2016.09.008>.
- Srinivasan, D., Choy, M.C., Cheu, R.L., 2006. Neural networks for real-time traffic signal control. *IEEE Trans. Intell. Transport. Syst.* 7 (3), 261–272.
- Sudheer, K., Nayak, P., Ramasastri, K., 2003. Improving peak flow estimates in artificial neural network river flow models. *Hydrol. Process.* 17 (3), 677–686.
- Tokar, A.S., Johnson, P.A., 1999. Rainfall-runoff modeling using artificial neural networks. *J. Hydrol. Eng.* 4 (3), 232–239.
- Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.* 239 (1), 132–147. [https://doi.org/10.1016/S0022-1694\(00\)00344-9](https://doi.org/10.1016/S0022-1694(00)00344-9).
- Veith, T.L., Sharpley, A.N., Arnold, J.G., 2008. Modeling a small, northeastern watershed with detailed, field-level data. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 51 (2), 471–483.

- Wagena, M.B., Bhatt, G., Buell, E., Sommerlot, A.R., Fuka, D.R., Easton, Z.M., 2019. Quantifying model uncertainty using Bayesian multi-model ensembles. *Environ. Model. Software* 117, 89–99. <https://doi.org/10.1016/j.envsoft.2019.03.013>.
- Wagena, M.B., Collick, A.S., Ross, A.C., Najjar, R.G., Rau, B., Sommerlot, A.R., Fuka, D. R., Kleinman, P.J.A., Easton, Z.M., 2018. Impact of climate change and climate anomalies on hydrologic and biogeochemical processes in an agricultural catchment of the Chesapeake Bay watershed, USA. *Sci. Total Environ.* 637–638, 1443–1454. <https://doi.org/10.1016/j.scitotenv.2018.05.116>.
- Wagena, M.B., Easton, Z.M., 2018. Agricultural conservation practices can help mitigate the impact of climate change. *Sci. Total Environ.* 635, 132–143. <https://doi.org/10.1016/j.scitotenv.2018.04.110>.
- Wang, E., Zhang, Y., Luo, J., Chiew, F.H., Wang, Q., 2011. Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. *Water Resour. Res.* 47 (5).
- Weigel, A.P., Liniger, M.A., Appenzeller, C., 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q. J. R. Meteorol. Soc.* 134 (630), 241–260. <https://doi.org/10.1002/qj.210>.
- Yurekli, K., Kurunc, A., Ozturk, F., 2005. Application of linear stochastic models to monthly flow data of Kelkit Stream. *Ecol. Model.* 183 (1), 67–75. <https://doi.org/10.1016/j.ecolmodel.2004.08.001>.
- Zealand, C.M., Burn, D.H., Simonovic, S.P., 1999. Short term streamflow forecasting using artificial neural networks. *J. Hydrol.* 214 (1), 32–48. [https://doi.org/10.1016/S0022-1694\(98\)00242-X](https://doi.org/10.1016/S0022-1694(98)00242-X).
- Zhang, G.P., Patuwo, B.E., Hu, M.Y., 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Comput. Oper. Res.* 28 (4), 381–396. [https://doi.org/10.1016/S0305-0548\(99\)00123-9](https://doi.org/10.1016/S0305-0548(99)00123-9).