



Hydrological modeling in a basin of the Brazilian Cerrado biome

ARTICLES doi:10.4136/ambi-agua.2639

Received: 03 Sep. 2020; Accepted: 18 Dec. 2020

Jéssica Assaid Martins Rodrigues^{1*} ; Alberto Carlos de Oliveira Andrade¹ ;
Marcelo Ribeiro Viola¹ ; Danton Diego Ferreira² ;
Carlos Rogério de Mello¹ ; Michael Silveira Thebaldi¹ 

¹Departamento de Recursos Hídricos. Universidade Federal de Lavras (UFLA), Caixa Postal 3037, CEP: 37200-900, Lavras, MG, Brazil. E-mail: alberto.ufla@gmail.com, marcelo.viola@ufla.br, crmello@ufla.br, michael.thebaldi@ufla.br

²Departamento de Automática. Universidade Federal de Lavras (UFLA), Caixa Postal 3037, CEP: 37200-900, Lavras, MG, Brazil. E-mail: danton@ufla.br

*Corresponding author. E-mail: je_assaid@yahoo.com.br

ABSTRACT

The Brazilian Cerrado biome (BCB) is among 25 biodiversity hotspots identified worldwide, and covers the recharge area of important aquifers and rivers in South America. The increase in deforestation has been threatening water availability in this region. In order to assist in the water-resource management of the BCB, this study models the daily streamflow in a basin of the Cerrado, using two approaches: a process-based model (Soil and Water Assessment Tool - SWAT) and the data-driven model (Artificial Neural Network - ANN). The performance of the models was evaluated by the Nash-Sutcliffe coefficient (NSE), coefficient of determination (R^2) and flow-duration-curves (FDC). The results indicate that SWAT (NSE > 0.61; R^2 > 0.68) and ANN (NSE > 0.91; R^2 > 0.79) models are suitable tools in daily streamflow modeling of the studied basin, with the ANN model being the most accurate. Based on FDC, the ANN model was also better than the SWAT model for all frequencies evaluated. Thus, the ANN model is a promising new approach for daily streamflow modelling in this region. Moreover, the results of this study can help water-resource managers in planning and implementing appropriate water allocation and conservation measures in the Brazilian Cerrado biome.

Keywords: brazilian savannah, hydrologic model, water resources.

Modelagem hidrológica em uma bacia do bioma Cerrado brasileiro

RESUMO

O bioma Cerrado brasileiro (BCB) está entre os 25 *hotspots* de biodiversidade identificados em todo o mundo, e abrange a área de recarga de importantes aquíferos e rios da América do Sul. O aumento do desmatamento vem ameaçando a disponibilidade de água na região. A fim de auxiliar na gestão dos recursos hídricos da BCB, este estudo tem como objetivo modelar a vazão diária em uma bacia do Cerrado, utilizando duas abordagens: um modelo baseado em processo (Soil and Water Assessment Tool - SWAT) e um modelo orientado por dados (Artificial Neural Network - ANN). O desempenho dos modelos foi avaliado pelo coeficiente de Nash-Sutcliffe (NSE), coeficiente de determinação (R^2) e curvas de duração de



fluxo (FDC). Os resultados indicam que os modelos SWAT ($NSE > 0,61$; $R^2 > 0,68$) e ANN ($NSE > 0,91$; $R^2 > 0,79$) são ferramentas adequadas na modelagem diária de vazões da bacia estudada, sendo o modelo ANN o mais preciso. Com base no FDC, o modelo ANN também foi melhor do que o modelo SWAT para todas as frequências avaliadas. Assim, o modelo ANN é uma nova abordagem promissora para modelagem de fluxo diário nesta região. Além disso, os resultados deste estudo podem ajudar os gestores de recursos hídricos no planejamento e implementação de medidas adequadas de alocação e conservação de água no bioma Cerrado brasileiro.

Palavras-chave: modelo hidrológico, recursos hídricos, savana brasileira.

1. INTRODUCTION

The Brazilian Cerrado biome is one of the most important areas in South America, covering the recharge areas of important aquifers and rivers, and approximately 204 million ha, which corresponds to 24% of Brazil (Medrado and Lima, 2014); it is recognized as the "cradle of Brazil's water" (Lima, 2011). Besides that, the Brazilian Cerrado biome is one of 25 biodiversity hotspots in the world, with high biological and endemic diversity, has suffered loss of vegetation due to agriculture and pasture expansion (Silva *et al.* 2006), which has resulting in a large number of endangered species (Myers *et al.*, 2000; Rodrigues *et al.*, 2020). These agriculture and pasture expansions also threaten the stream flows from watersheds (Silva and Bates, 2002); however, the understanding of their impacts on streamflow in the Cerrado biome is still limited (Beuchle *et al.* 2015). Improving the knowledge base on hydrological modeling in the Cerrado biome is therefore important for water-resource management, since it allows the quantification of current and future water availability, which is essential to ensure water security and economic development.

Hydrological models provide a representation of the processes in the hydrological cycle of a basin and help in understanding, predicting and managing water resources (Devia *et al.*, 2015). Among the various hydrological models that have been developed, the Soil and Water Assessment Tools (SWAT) model is one of the most applied for simulating at basin-scale around the world. SWAT is a semi-conceptual and semi-distributed hydrological model developed by the Agricultural Research Service (ARS/USA) and Texas A&M University in the early 1990s (Arnold *et al.*, 1998). The SWAT has been used to predict streamflow time series, hydrological processes, water balance, and for evaluating the impacts of climate change, land-use change, and different management practices on the surface hydrological cycle, sediment yield, water quality (Abbaspour *et al.* 2007). Many researchers have applied the SWAT model for different basins. Pontes *et al.* (2016) applied the SWAT hydrological model to estimate daily and monthly discharges for the Camanducaia River Basin, Brazil, and to evaluate the performance of these calibrations in a contiguous drainage basin. Monteiro *et al.* (2015) used two precipitation grids, CFSR (Climate Forecast System Reanalysis) and WFDEI (WATCH Forcing Data methodology applied to ERA-Interim), as inputs to a SWAT model for river discharge simulation in the Tocantins catchment, Brazil. Choubin *et al.* (2019) compared the SWAT and Identification of Hydrographs and Components from Rainfall, Evaporation, and Stream (IHACRES) models in the streamflow regionalization for the Karkheh River Basin, Iran. Rodrigues *et al.* (2020) applied a SWAT model to simulate the monthly streamflow for three basins of the Brazilian Cerrado biome. Alvarenga *et al.* (2020) compared the SWAT and Variable Infiltration Capacity (VIC) models in the monthly streamflow simulation for the Verde River Watershed, located in the Minas Gerais state in southern Brazil. However, despite the wide use of the SWAT model, it requires a large amount of temporal and spatial data, maps, and input parameters, that are sometimes hard to predict (Makwana and Tiwari, 2014).

Therefore, it is not applicable (or presents a worse performance) to basins in which input data for hydrological modeling are not available (Yaseen *et al.*, 2019). This is the case of most of the Brazilian Cerrado biome basins where there is a significant lack of data (Nóbrega *et al.*, 2017).

An alternative to conceptual models is the Artificial Neural Network (ANN), which is an empirical model capable of learning nonlinear relationships between the variables of a process, and relating inputs and outputs without the need of a detailed understanding of its physical characteristics. This feature makes ANN an effective tool for modeling complex hydrological processes (Talebizadeh and Moridnejad, 2011). Thus, ANN has been widely applied to solve several water-resource problems, and it was found to be a powerful tool for streamflow simulation (Aichouri *et al.*, 2015). Kothari and Gharde (2015) applied ANN for the streamflow modeling of Savitri catchment, India. Zhou *et al.* (2018) forecasted the monthly streamflow of the Jinsha River by using three (ANN) architectures: extreme learning machine, radial basis function network, and Elman network. Vilanova *et al.* (2019) applied ANN to simulate daily streamflows for Brazilian Atlantic Rainforest basins. Papalaskaris (2020) applied ANN for the daily low streamflows forecast of Iokastis Stream, Kavala City, NE Greece, NE Mediterranean Basin. However, no study has applied the ANN model to simulate streamflow in basins of the Brazilian Cerrado biome.

The implementation of models with different approaches to the same problem allows exploring the advantages and disadvantages of the models and finding the best and most efficient structure for a given region. Few studies have compared SWAT and ANN models for streamflow estimation (Tan *et al.*, 2020). Demirel *et al.* (2009) assessed the performance of SWAT and ANN models for the daily flow forecasting in the Pracana Basin, Portugal, and determined that the ANN model yielded the highest accuracy ratio. Noori and Kalin (2016) assessed the performance of SWAT and ANN models for the daily runoff simulation in the Gwinnett and Atlanta watersheds, and determined that the ANN model showed the better performance. Jimeno-Saiez *et al.* (2018) assessed the performance of SWAT and ANN models for daily streamflows simulation in different climatic zones of Peninsular Spain, and reported that ANN model had performed better for maximum values and for minimum values the SWAT model performance had been better. To the best of our knowledge, there is no study regarding streamflow estimation by comparing the SWAT and ANN in a Brazilian basin. Thus, such assessment is still limited in the literature, and therefore requires more investigation.

In this context, in order to assist in the water-resource management of the Cerrado biome, this study assesses the suitability of the SWAT and ANN models to simulate the daily streamflow in the Manuel Alves da Natividade River Basin (MRB), located in the Brazilian Cerrado biome, and compares their performances to determine which is more appropriate for the studied basin. The MRB was selected based on the great importance that it represents, since it serves several hydroelectric projects, and one of the major irrigation projects in Brazil, the Manuel Alves irrigation project, is installed on it (Tocantins, 2012). This study will provide a basis for the use of local and public administrations in improving a successful basin management strategy in the MRB. It will also help spread the research to different regions of the world.

2. MATERIALS AND METHODS

2.1. Study area

The Manuel Alves da Natividade River Basin (MRB) has a drainage area of 14,344 km² (Figure 1) and is one of the main sub-basins of the Tocantins-Araguaia River Basin (TARB), which in turn is the largest basin entirely inserted in the Brazilian territory. The MRB is fully inserted in the Cerrado biome, and the streamflow produced by it directly feeds the Luís Eduardo Magalhães hydropower plant, with an installed capacity of 903 MW (ONS, 2020),

corresponding to 25% of its incremental contributing area. It also feeds the Manuel Alves irrigation project, with an irrigable area of 20 thousand hectares, which is one of the major irrigation projects in Brazil (Tocantins, 2012). The MRB is located in southeastern Tocantins state, north Brazil, between latitude 11°09'45" to 12°14'54" S and longitude 46°33'04" to 48°18'40" W, and is limited by the basins of the Rivers Palma from south, Balsas from north, Tocantins from west and São Francisco from east.

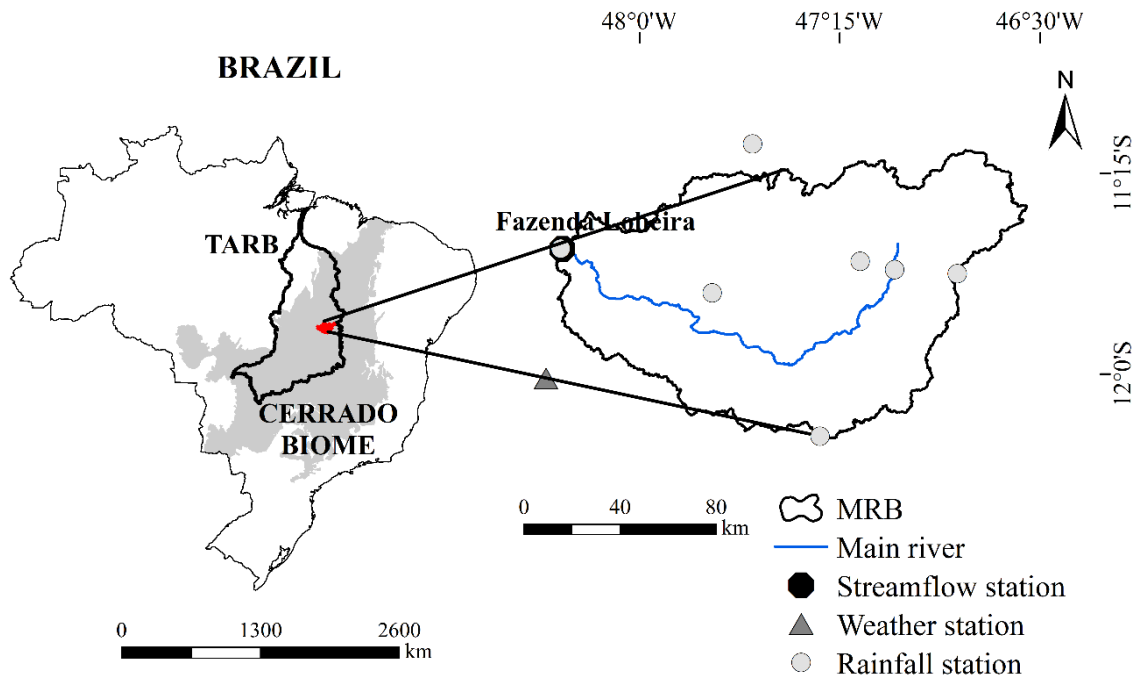


Figure 1. Location of the Manuel Alves da Natividade River Basin and monitoring stations.

According to the Köppen classification, the basin climate is Aw (tropical savanna), characterized by a rainy summer and dry winter (Kottek *et al.*, 2006). The average annual precipitation is 1500 mm and the average annual temperature is 25°C.

2.2. Soil and Water Assessment Tool (SWAT)

The SWAT is a physically based semi-distributed hydrological model, free of charge, available at (<http://swat.tamu.edu/>), and widely used across the world. To execute the model, various inputs are required, such as hydrometeorological data, soil and land-use maps and a digital elevation model (DEM) in the basin. The SWAT model divides the basin into multiple sub-basins based on the river network and topography. Each sub-basin is further divided into Hydrologic Response Units (HRU), which consist of unique combinations of soil class, land use and slope (Arnold *et al.*, 1998). SWAT simulations are based on water balance, are performed for each HRU and are accumulated to obtain the total for the sub-basins and basin. Equation 1 describes the water balance adopted by the SWAT model.

$$SW_t = SW_0 + \sum_{i=1}^n (R_{\text{day}} - Q_{\text{surf}} - E_a - W_{\text{seep}} - Q_{\text{gw}}) \quad (1)$$

Where SW_t is the final amount of soil water for the day (mm), SW_0 is the initial amount of soil water for the day (mm), R_{day} is the total rainfall for the day (mm), Q_{surf} is the surface runoff for the day (mm), E_a is the evapotranspiration for the day (mm), W_{seep} is the total amount of water that seeps through the base of the soil profile for the day (mm), and Q_{gw} is the groundwater flow for the day (mm). For a detailed description of the SWAT model, refer to Arnold *et al.* (1998) and Neitsch *et al.* (2011).

To perform the hydrological simulation, SWAT requires hydrometeorological data (rainfall, maximum and minimum air temperatures, solar radiation, relative humidity, and wind speed), and geospatial data, which include the digital elevation model (DEM), land-use map and soil map. For the MRB, daily hydrometeorological data were obtained from seven rainfall gauge stations and one weather station (Figure 1) in the Brazilian National Water Agency (ANA-Hidroweb) and the Meteorological Database for Education and Research (BDMEP) of the Brazilian National Institute of Meteorology (INMET). The ASTER DEM of the MRB (Figure 2a), with spatial resolution of 30m, was obtained from the USGS (United States Geological Survey), and showed elevations ranging from 197 to 960 m. The land-use map of the MRB (Figure 2b) was provided by the Tocantins Planning and Budget Secretariat (SEPLAN) (Tocantins, 2012), and its land-use classes consist of Cerrado (79.24%), Riparian Forest (8,38%), Pasture (6.56%), Amazon Forest (5.56%), Water bodies (0.18%), Urban Area (0.04%), Agriculture (0.02%) and bare soils (0.02%). The soil map of the MRB (Figure 2c) was obtained from the Brazilian Agricultural Research Corporation (EMBRAPA, 2011) at a 1:5,000,000 scale. The soil classes in the MRB are Red-Yellow Argisol (PVA; 52.85%), Litholic Neosol (RL; 14.08%), Red-Yellow Latosol (RYL; 10.69%), Endopetric Plinthosol (FF; 8.46%), Haplic Cambisol (CX; 5.04%), Quartzarenic Neosol (RQ; 4.73%), Yellow Latosol (LA; 2.16%), Haplic Gleysol (GX; 1.24%) and Haplic Plinthosol (FX; 0.75%).

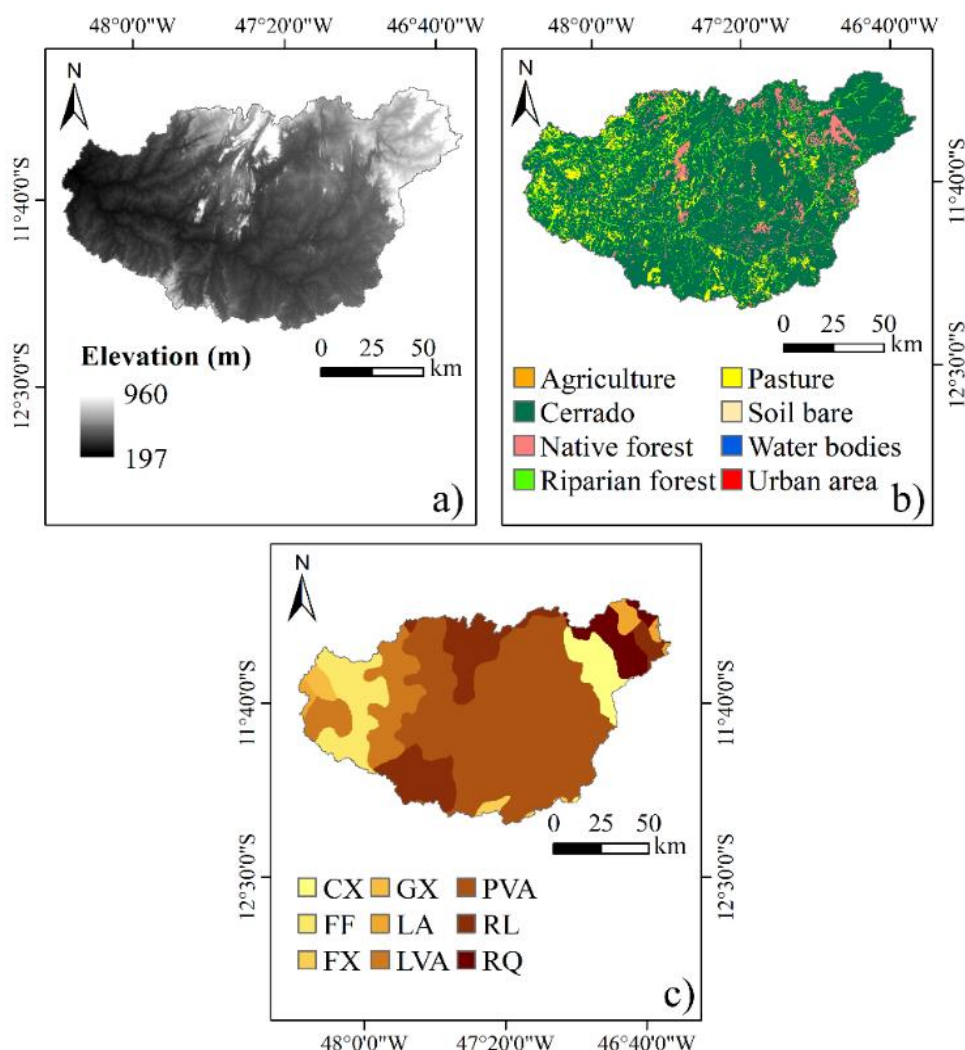


Figure 2. Digital elevation model (a), land use map (b) and soil map (c) of the Manuel Alves da Natividade River Basin.

For the streamflow modeling, the SWAT model was set up for the period between 1983 and 2005 with daily precipitation, maximum and minimum temperature, solar radiation, wind speed and relative humidity data. The first three years were used as a warm-up period to reduce the uncertainties regarding the initial conditions of the surface domain (Mello *et al.*, 2008). The period from 1986 to 1995 was used for calibration, whereas the period from 1996 to 2005 was used for validation.

The parameters used in calibration (Table 1) were selected based on previous studies conducted in the Cerrado biome, such as Oliveira *et al.* (2019), Rodrigues *et al.* (2020) and Amorim *et al.* (2020), and based on Latin Hypercube One-factor- At-a-Time sensitivity analysis (LH-OAT) method (Van Griensven *et al.*, 2006). LH-OAT was applied using the SUFI2 algorithm in the SWAT-CUP software. Following sensitivity analysis, the daily streamflow calibration was performed manually by changing one parameter at a time, as performed by Nyeko (2015), Pereira *et al.* (2016) and Rodrigues *et al.* (2020), and the validation was performed updating the SWAT model with the parameters obtained in the calibration period.

Table 1. SWAT model parameters and their final calibrated values in the daily time step for the MRB.

Parameters	Description	Calibrated value
V_ESCO.hru	Soil water evaporation compensation coefficient	0.500
R_CN2.mgt	Initial curve-number for moisture conditions II	-0.040
V_ALPHA_BF.gw	Baseflow recession coefficient (days)	0.069
A_GW_DELAY.gw	Time interval for aquifer recharge (days)	42.00
V_CANMX.hru	Maximum canopy storage (mm)	28.00
V_CH_K2.rte	Effective hydraulic conductivity of the main channel (mm.h ⁻¹)	3.690
V_CH_N2.rte	Manning's number for the main channel	0.056
V_GW_REVAP.gw	Water seepage coefficient for the unsaturated zone	0.200
R_SOL_AWC ().sol	Soil water storage capacity (mm _{water} .mm _{soil} ⁻¹)	0.230
R_SOL_K (). Sol	Saturated hydraulic conductivity (mm.h ⁻¹)	-0.072
V_CH_N1.sub	Manning's number for the secondary channel	0.520
V_CH_K1.sub	Effective hydraulic conductivity of the tributary channel (mm.h ⁻¹)	4.740
V_SLSOIL.hru	Slope length (m)	17.629
V_LAT_TTIME.htm	Lateral flow propagation time (days)	5.000

Note: Prefixes “V,” “R” and “A” correspond to the operations “replace by a given value,” “relative - an existing parameter value is multiplied by” and “add to the existing parameter value,” respectively.

2.3. Artificial Neural Network (ANN)

The ANN model is a data-driven mathematical model that was developed to imitate the structure of a human brain neural network and has been widely applied to solve water-resource problems (Minns and Hall, 1996). To execute the model, prior knowledge of the physical

characteristics of the process is not required, only hydrological and meteorological data from the basin. The ANN model uses interconnected artificial neurons that receive input information (x_1, x_2, \dots, x_n), perform operations and provide an output (y). Each connection between neurons (or synapse) has an associated intensity, expressed by a weight (w_1, w_2, \dots, w_n). Each neuron determines an input value (net) through the sum of the products of the input weighted by respective values. The weights are the values that represent the degree of importance of each input to the neuron, obtained at the time of neural network training. Once the weight is determined, it becomes the activation value of the respective neuron. This value is a function of the input ($y = f(\text{net})$). An activation function precedes the transfer function and has the task of passing the signal obtained through the inputs to the transfer function (Haykin, 2007).

Multilayer perceptron (MLPs) ANN model has hidden layers, and according to Haykin (2007) they have three main characteristics: a) the model of each neuron has a non-linear activation function; b) the network has at least one hidden layer; and c) the network has a high degree of connectivity between its processing elements.

In this study, a feedforward neural network was defined with three layers: input, internal (or hidden) and output layers. This type of network is widely used in signal filtering, data compression, pattern recognition and inter-comparison of patterns. For this neural network, the hyperbolic tangent activation function (tansig) was used for the neurons of the middle layer; the linear activation function (purelin) for the output neuron; and the Bayesian approximation algorithm (trainbr) for network training (Haykin, 2007).

For structuring the MLP artificial neural network, 17 input nodes were used, referring to rainfall, maximum temperature, minimum temperature, solar radiation, potential evaporation, mean temperature, relative humidity, wind speed and streamflow of the previous day. To avoid polarization of the neural network and delays in the learning process, the input variables were normalized between -1 and 1. For this purpose, each variable was divided by the absolute maximum value of the set of corresponding variables. Concerning the middle layer, the number of neurons ranged from 1 to 10, and 30 replicates were generated for each neuron. Because the goal of the network was to model the streamflow of the MRB, only one neuron was used in the output layer. Figure 3 shows the MLP artificial neural network architectures analyzed.

To generate the MLP, the MATLAB® software was used, and to compare the results of the various neural models, Tukey's test (TSD – Tukey's Significant Difference) was applied to determine whether there was a significant difference (5% significance level) between the mean Nash-Sutcliffe coefficients (NSE) (Tukey, 1949).

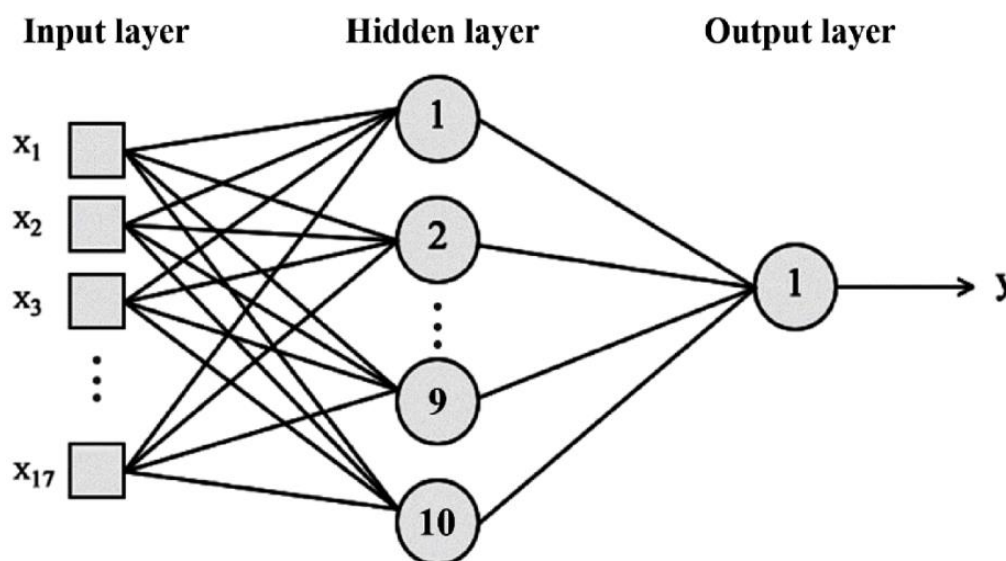


Figure 3. Multilayer perceptron neural network architectures analyzed in this study.

2.4. Model Performance Evaluation

In order to evaluate the performance of the SWAT and ANN models, both in the calibration/training and validation periods, the coefficient of determination (R^2) and the Nash-Sutcliffe coefficient (NSE) were used. R^2 describes the proportion of the variance in observed data that can be explained by the model. It ranges from 0 to 1, and the higher its value, the better the fit. R^2 is calculated by Equation 2 and, in general, values greater than 0.6 are considered acceptable (Moriassi *et al.*, 2015).

$$R^2 = \left[\frac{\sum_{i=1}^n (Q_{obsi} - \overline{Q_{obs}}) \cdot (Q_{simi} - \overline{Q_{sim}})}{\sqrt{\sum_{i=1}^n (Q_{obsi} - \overline{Q_{obs}})^2 \cdot \sum_{i=1}^n (Q_{simi} - \overline{Q_{sim}})^2}} \right]^2 \quad (2)$$

Where Q_{obs} is the observed streamflow, $m^3 \cdot s^{-1}$, Q_{sim} is the simulated streamflow, $m^3 \cdot s^{-1}$, $\overline{Q_{obs}}$ is the observed mean streamflow, $m^3 \cdot s^{-1}$, $\overline{Q_{sim}}$ is the simulated mean streamflow, $m^3 \cdot s^{-1}$, and n is the number of data points.

NSE indicates how well the plot of observed versus simulated data fits the 1:1 line. It ranges from $-\infty$ to 1, and the higher the value, the better the fit of the simulated to the observed data (Krause *et al.*, 2005). NSE is calculated using the following Equation 3:

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obsi} - Q_{simi})^2}{\sum_{i=1}^n (Q_{obsi} - \overline{Q_{obs}})^2} \quad (3)$$

Moriassi *et al.* (2015) proposed the following classification for NSE using a daily time step for simulations: $NSE > 0.8$, the model is considered very good; $0.7 < NSE < 0.8$, the model is considered good; and $0.5 < NSE < 0.7$, the model is considered satisfactory.

3. RESULTS AND DISCUSSION

3.1. Calibration and validation of the SWAT model

The SWAT model has been calibrated from 1986 to 1995 and validated from 1996 to 2005 using daily streamflow data from the Fazenda Lobeira gauging station, which delimits the MRB. The performance of the SWAT model in terms of R^2 was acceptable ($R^2 > 0.6$), with values of 0.70 and 0.68 for the calibration and validation periods, respectively. Regarding the NSE, values of 0.67 and 0.61 were obtained for the calibration and validation periods, respectively. These values allow classifying the model as “satisfactory” in both periods (Moriassi *et al.* 2015). These results demonstrate that the model is able to satisfactorily simulate the observed daily streamflow. Studies on the application of SWAT in different basins in Brazil have used NSE to evaluate the performance of the simulations. Durães *et al.* (2011) evaluated the performance of the SWAT model in hydrological simulation of the Paraopeba River Basin, with a 10,200 km² drainage area, and obtained NSE value of 0.79 for both the calibration and validation periods. Pereira *et al.* (2016) evaluated the performance of the SWAT model in hydrologic simulations of the Pomba River Basin, with a drainage area of 8,600 km², and obtained NSE of 0.76 in both the calibration and validation periods. Rodrigues *et al.* (2020) evaluated the performance of the SWAT model in monthly streamflow simulation of three hydrographic basins located in the Cerrado biome, and obtained NSE values ranging from 0.56 to 0.84 and 0.70 to 0.81, respectively, for calibration and validation periods.

Table 1 presents the parameters that were used in the SWAT model, which were selected from the literature review and the LH-OAT sensitivity analysis method, and the final calibrated values of the best simulation generated in calibration for MRB. A detailed description of the

parameters used in this study can be obtained in Neitsch *et al.* (2011).

3.2. Training and validation of ANN model

The ANN model was developed using the same hydrometeorological dataset and the same training/calibration (1986 to 1995) and validation (1996 to 2005) periods used in the SWAT model. The training and validation of the ANN model was carried out with the number of neurons in the middle layer (NNML) between 1 and 10, and the NSE results for different ANN architectures are presented in Table 2. Considering the classification of Moriasi *et al.* (2015) as reference, the NSE values for the ANNs are classified as “very good” (>0.8). The results showed an increase in the quality of the simulation when increasing the NNML up to 3. However, above this threshold, no improvement was observed in the NSE values with increasing the NNML.

Table 2. NSE of ANN model with the NNML between 1 and 10.

NNML	1	2	3	4	5	6	7	8	9	10
Training	0.90	0.90	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
Validation	0.89	0.90	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91

Figure 4 shows Tukey’s statistical test results at the 5% significance level for the mean NSE obtained from the 30 replicates for each ANN architecture. Taking the result of the network with a single neuron in the middle layer as reference (Figure 4a), a significantly different behavior of the ANNs with a higher NNML is observed. For the network with the NNML set to 2 (Figure 4b), the behavior was statistically similar to that of the networks with the NNML set to 3 and 4, and distinct from the others. The networks with the NNML set to 3 and 4 were significantly different from the network with the NNML set to 1. For networks with the NNML ≥ 5 , there was an analogous behavior that was significantly different from networks with the NNML set to 1 and 2. Thus, the structure of 5 NNML was chosen as the proper structure in the Manuel Alves da Natividade River Basin and for the remainder of the analysis, since it was the least complex among those with the best results. The R^2 values (for 5 NNML) were 0.97 and 0.79 for the training and validation periods, respectively.

3.3. Comparison of model performance

The comparison of the process-based SWAT model and the data-driven ANN model was carried out to study the suitability of these models to simulate the daily streamflow in a basin of the Brazilian Cerrado biome. Table 3 shows the comparison of SWAT and ANN models for the calibration/training and validation periods. The results suggest that the ANN model had greater performance (higher NSE and R^2) than SWAT for the entire simulation period. This result is similar to those obtained by Makwana and Tiwari (2017), Jimeno-Sáez *et al.* (2018) and Ahmadi *et al.* (2019), which observed a better performance of the ANN model than the SWAT model, considering both R^2 and NSE coefficients.

SWAT and ANN performances can also be observed by means of scatter plots (Figure 5a, b), for the calibration/training and validation periods, respectively. From these graphics, it can be observed that the scattered SWAT points are concentrated above the 1: 1 line (perfect fit), which means a tendency to overestimate the streamflow. In addition, the ANN model presented better estimates for almost all the streamflow values. In general, the scattered ANN points are closer to the 1:1 line (perfect fit) than SWAT points, which means greater precision.

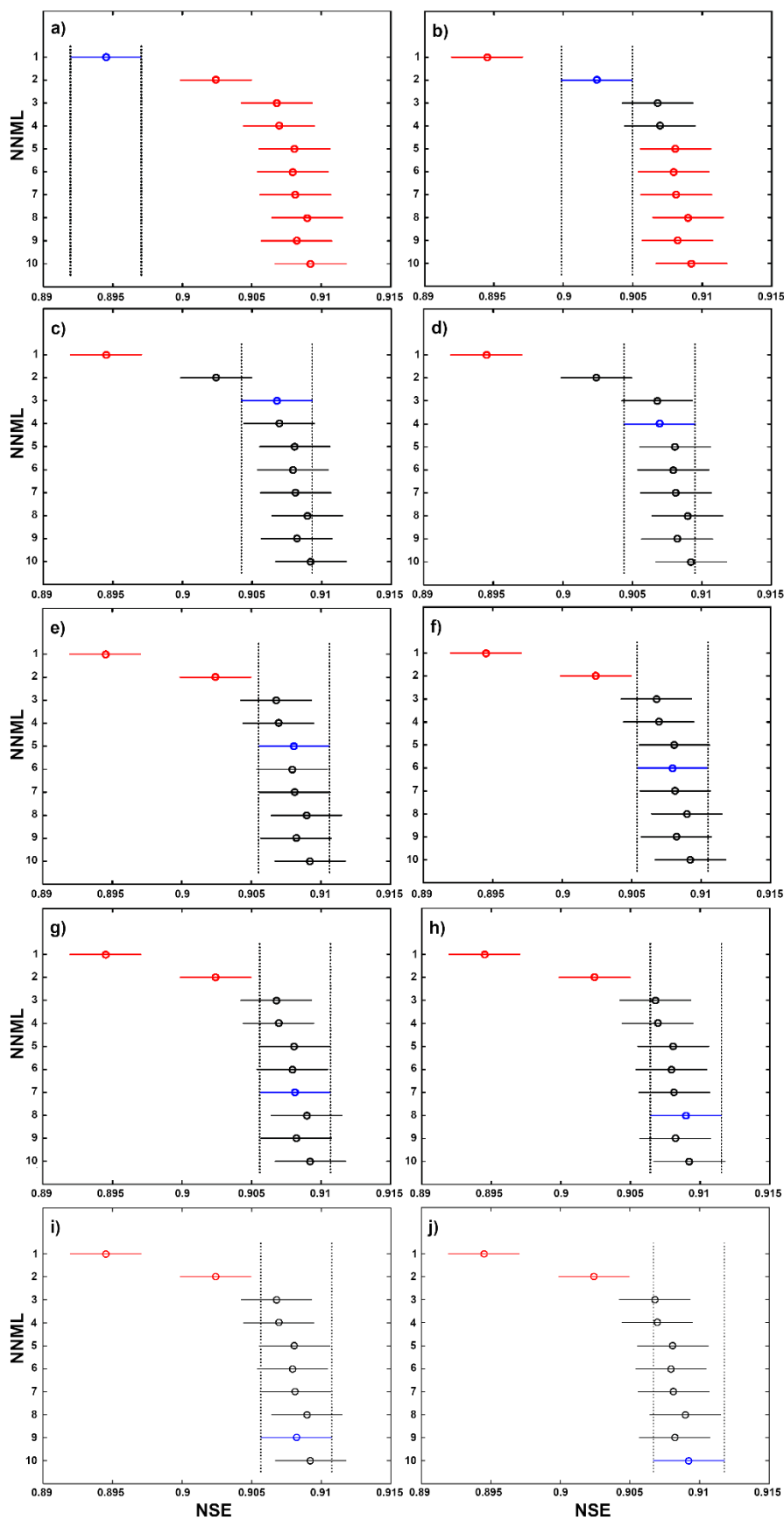


Figure 4. Tukey's test at the 5% probability significance level for the mean NSE for ANN configurations with 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), 6 (f), 7 (g), 8 (h), 9 (i) and 10 (j) neurons in the middle layer.

Table 3. Performances of SWAT and ANN models in the calibration/training and validation periods with a daily time step.

Model	Calibration/Training		Validation	
	NSE	R ²	NSE	R ²
SWAT	0.67	0.70	0.61	0.68
ANN	0.91	0.97	0.91	0.79

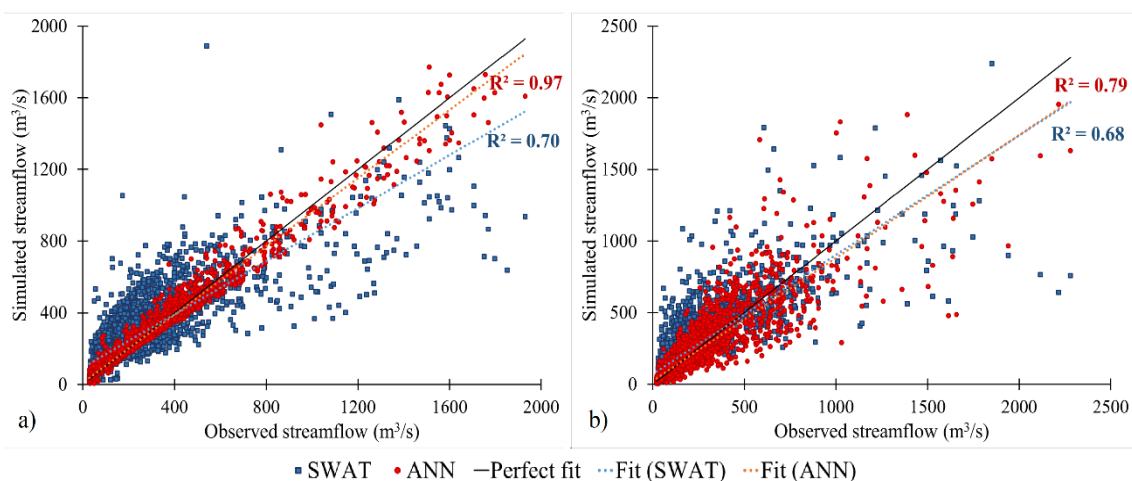
**Figure 5.** Scatter plots between the observed daily streamflow and the results of SWAT and ANN models during (a) calibration/training period and (b) validation period.

Figure 6 shows the observed daily rainfall and streamflow data along with the streamflow data simulated by the ANN (a) and the SWAT (b) models for the calibration and validation periods. The results indicate that both models showed a good performance when compared against the observed streamflow. However, the adherence of the simulated to the observed data shows an underestimation of the maximum streamflow values by both models. The peak discharges are related to the surface runoff as a response to the intense rainfall events. However, the magnitude of the maximum streamflow depends not only on the magnitude of the intense rainfall but also on other factors, such as soil moisture, topography and land-use, making the modeling of maximum events a hard task in the context of continuous simulations. Otherwise, for recession periods, a greater agreement between the curves was observed. The greater performance of the models in the recession periods is explained based on the predominance of groundwater runoff, which is easier to model than direct surface runoff, given that its genesis is related to the discharge of the aquifer, following the Darcy's law for fluidity in porous media.

In order to further assess the difference between the SWAT and ANN models in the streamflow simulation in the MRB, the flow duration curves (FDC) were developed (Figure 7). It can be observed that the simulated curves presented good agreement with the observed one. The ANN model performed better in all streamflow simulations, and the SWAT model performed worse in estimating intermediate streamflow-frequency values.

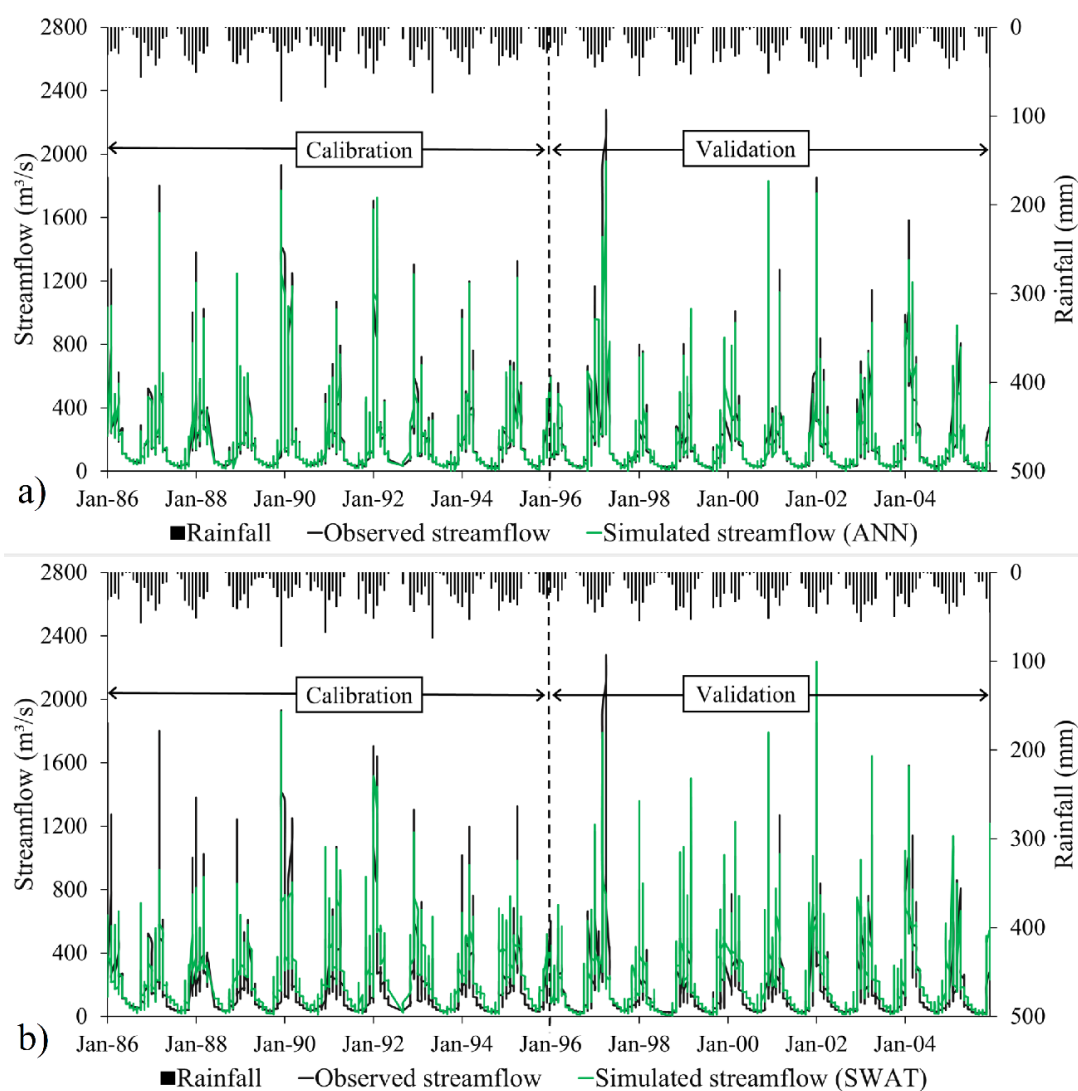


Figure 6. Observed and simulated hydrographs (ANN (a) and SWAT (b)) during the calibration and validation periods for MBR.

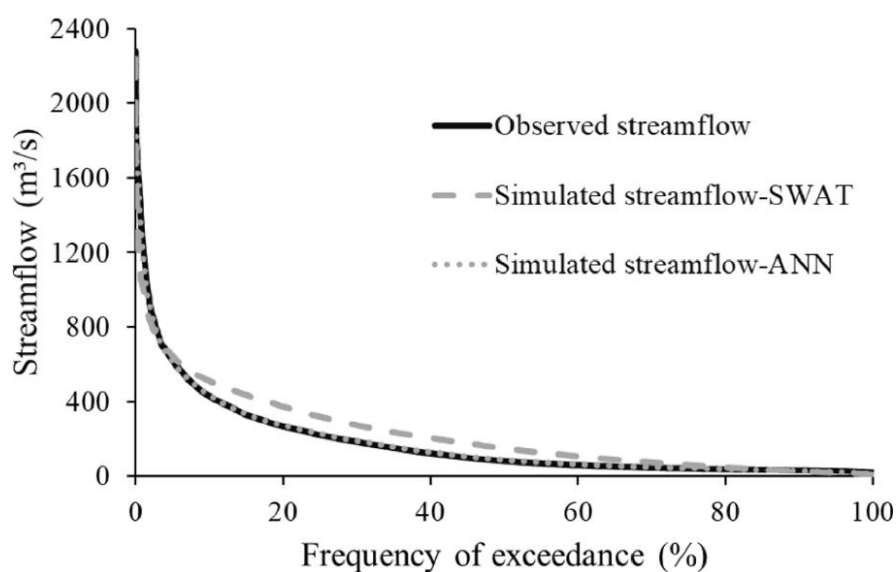


Figure 7. Exceedance frequency curves of observed streamflow, and of simulated SWAT and ANN streamflows.

Table 4 shows the percentage difference of simulated streamflows with different frequencies of exceedance by the SWAT and ANN models compared to the observed streamflows. The best results were obtained from the ANN model, so its error rates were lower than those of the SWAT model for all simulated streamflows. The ANN model had a very good performance in estimating maximum values and a good performance in estimating minimum values, with error rates ranging from 0.5 to 20.3%. In turn, the SWAT model had a good performance in estimating maximum values, but in estimating average values, its performance was poor, showing error rates ranging from 4.5 to 83.1%. When similar studies in the literature are examined, results similar to this study are obtained. Singh (2016), evaluating the ANN and SWAT models to simulate the streamflow for an Agricultural Watershed in India, reported that the ANN model estimates the streamflow values more accurately and with less uncertainty. Koycegiz and Buyukyildiz (2019), evaluating the ANN and SWAT models to simulate the streamflow at the headwater of Çarsamba River, located at the Konya Closed Basin, Turkey, reported that the ANN model was more successful than the SWAT model. Ahmadi *et al.* (2019), evaluating the ANN and SWAT models to simulate the daily, monthly, and annual streamflows for the Kan Watershed, Iran, reported that the ANN model performed better in all streamflow simulations. However, when evaluating the literature, studies with different results are also found. Zakizadeh *et al.* (2020) analyzed the performance of the SWAT and ANN models in the runoff simulation for the Darake Watershed, Iran. They found that, for the maximum values, the performance of ANN model was better than that of SWAT, and for the minimum values the performance of SWAT model was better than that of ANN. Kim *et al.* (2015) analyzed the SWAT and ANN models for streamflow estimation of the Samho gauging station at Taehwa River, Korea. They found that the ANN was better at estimating high flows, while the SWAT model was better at simulating low flows. Pradhan *et al.* (2020) analyzed the SWAT and ANN models to simulate daily streamflow in three different river basins in different climatic regions of Asia. They found that in general, in two basins, the performance of the ANN model is better than the SWAT, whereas, in the other basin, the SWAT model performance is better than the ANN. Also, the SWAT model was found to be better for low flow simulation and the ANN model performed better for high flow simulation in the three river basins. Jimeno-Sáez *et al.* (2018) analyzed the SWAT and ANN models to simulate streamflow in two catchments in Peninsular Spain. They reported that the SWAT was more successful in relation to better simulation of lower flows, while ANN was superior at estimating higher flows in all cases. Thus, it can be seen that there is not a better model for all regions; thereby, one can highlight the importance of this analysis for the proper management of water resources.

Table 4. Observed streamflows, in $\text{m}^3 \text{s}^{-1}$, with exceeding of 5% ($Q_{5\%}$), 10% ($Q_{10\%}$), 20% ($Q_{20\%}$), 50% ($Q_{50\%}$), 90% ($Q_{90\%}$) and 95% ($Q_{95\%}$), and the percentage differences (ΔQ) of these simulated streamflows by the SWAT and ANN model in comparison against the observed streamflow.

Percentage exceedance	$Q_{\text{observed}} (\text{m}^3 \text{s}^{-1})$	$\Delta Q - \text{SWAT} (\%)$	$\Delta Q - \text{ANN} (\%)$
$Q_{5\%}$	618.7	4.5	-0.5
$Q_{10\%}$	430.0	19.1	1.0
$Q_{20\%}$	265.9	40.2	2.8
$Q_{50\%}$	81.9	83.1	7.0
$Q_{90\%}$	32.6	-14.1	-10.6
$Q_{95\%}$	29.2	-36.0	-20.3

Therefore, our results suggest that the use of ANN and SWAT models is suitable for simulating the daily streamflow in the MRB, with the ANN model showing better performance. However, although the ANN model showed better streamflow simulation ability than SWAT,

ANN model is a data-driven model, which does not consider the hydrological processes involved. Therefore, it is not recommended for situations in which there are alterations of the physical characteristics of the basin, like simulating hydrologic impacts under climate or land use change scenarios, since the rainfall network and land-use types are changed. In contrast, the process-based SWAT model describes the rainfall-streamflow transformation processes in detail; thus, it can better simulate the response of streamflow to changes in environmental factors than the data-driven model.

4. CONCLUSIONS

The objective of this study was to model the daily streamflow in a basin of the Brazilian Cerrado biome, using two approaches: a process-based model (the Soil and Water Assessment Tool - SWAT) and the data-driven ANN model, comparing their performance to determine which is more appropriate for the studied basin. It was concluded that both SWAT and ANN were suitable to simulate daily streamflow in the MRB. However, ANN model was better than the SWAT model during calibration and validation periods in all evaluations performed.

The multilayer perceptron artificial neural network was used to model daily streamflow in the MRB based on rainfall, streamflow and meteorological data. The best configuration had 5 neurons in the middle layer. The results of the statistical coefficients showed very good performance in the daily streamflow modeling. In addition, one of the advantages of the ANN model is that it does not require any physical characteristics of the basin and, therefore, its implementation is easier. However, it does not consider hydrological processes, thus, the ANN model cannot be used to simulate the streamflow if the physical characteristics in the basin change. Therefore, ANN is recommended for situations requiring real-time streamflow predictions, as is the case for civil surveillance.

The SWAT hydrological model produced adequate results for daily streamflow modeling in the MRB. It is concluded that the greater complexity of this method, the need for hydrological experience and the higher computational demand are justified for situations in which the analysis of processes of the hydrological cycle are important in a particular drainage basin, as for example in the simulation of land-use changes or climate changes scenarios.

Therefore, the results of this study can help in understanding the capabilities of both models to simulate daily streamflow in the MRB, and contribute to the proper management of water resources in the Brazilian Cerrado biome.

5. REFERENCES

- ABBASPOUR, K. C.; YANG, J.; MAXIMOV, I.; SIBER, R.; BOGNER, K.; MIELEITNER, J. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. **Journal of Hydrology**, v. 333, n. 2-4, p. 413-430, 2007. <https://doi.org/10.1016/j.jhydrol.2006.09.014>
- AHMADI, M.; MOEINI, A.; AHMADI, H.; MOTAMEDVAZIRI, B.; ZEHTABIYAN, G. R. Comparison of the performance of SWAT, IHACRES and artificial neural networks models in rainfall-runoff simulation (case study: Kan watershed, Iran). **Physics and Chemistry of the Earth, Parts A/B/C**, v. 111, p. 65-77, 2019. <https://doi.org/10.1016/j.pce.2019.05.002>
- AICHOURI, I.; HANI, A.; BOUGHERIRA, N.; DJABRI, L.; CHAFFAI, H.; LALLAHEM, S. River flow model using artificial neural networks. **Energy Procedia**, v. 74, p. 1007-1014, 2015. <https://doi.org/10.1016/j.egypro.2015.07.832>

- ALVARENGA, L. A.; CARVALHO, V. S. O.; OLIVEIRA, V. A. D.; MELLO, C. R. D.; COLOMBO, A.; TOMASELLA, J.; MELO, P. A. Hydrological simulation with SWAT and VIC Models in the Verde River Watershed, Minas Gerais. **Revista Ambiente & Água**, v. 15, n. 4, 2020. <https://doi.org/10.4136/ambi-agua.2492>
- AMORIM, J. D. S.; VIOLA, M. R.; JUNQUEIRA, R.; OLIVEIRA, V. A. D.; MELLO, C. R. D. Evaluation of Satellite Precipitation Products for Hydrological Modeling in the Brazilian Cerrado Biome. **Water**, v. 12, n. 9, p. 2571, 2020. <https://doi.org/10.3390/w12092571>
- ARNOLD, J. G.; SRINIVASAN, R.; MUTTIAH, R. S.; WILLIAMS, J. R. Large area hydrologic modeling and assessment part I: model development 1. **JAWRA Journal of the American Water Resources Association**, v. 34, n. 1, p. 73-89, 1998. <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>
- BEUCHLE, R.; GRECCHI, R. C.; SHIMABUKURO, Y. E.; SELIGER, R.; EVA, H. D.; SANO, E.; ACHARD, F. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. **Applied Geography**, v. 58, n. 2, p. 116–127, 2015. <https://dx.doi.org/10.1016/j.apgeog.2015.01.017>
- CHOUBIN, B.; SOLAIMANI, K.; REZANEZHAD, F.; ROSHAN, M. H.; MALEKIAN, A.; SHAMSHIRBAND, S. Streamflow regionalization using a similarity approach in ungauged basins: Application of the geo-environmental signatures in the Karkheh River Basin, Iran. **Catena**, v. 182, p. 104128, 2019. <https://doi.org/10.1016/j.catena.2019.104128>
- DEMIREL, M. C.; VENANCIO, A.; KAHYA, E. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. **Advances in Engineering Software**, v. 40, n. 7, p. 467-473, 2009. <https://doi.org/10.1016/j.advengsoft.2008.08.002>
- DEVIA, G. K.; GANASRI, B. P.; DWARAKISH, G. S. A review on hydrological models. **Aquatic Procedia**, v. 4, p. 1001-1007, 2015. <https://doi.org/10.1016/j.aqpro.2015.02.126>
- DURÃES, M. F.; MELLO, C. R.; NAGHETTINI, M. Applicability of the SWAT model for hydrologic simulation in Paraopeba River Basin, MG. **Cerne**, v. 17, n. 4, p. 481-488, 2011.
- EMBRAPA. **Mapa de solos do Brasil**. 2011. Available at: http://mapoteca.cnps.embrapa.br/geoacervo/det_mapa.aspx. Access: 15th March 2017.
- HAYKIN, S. **Redes neurais: princípios e prática**. São Paulo: Bookman, 2007.
- JIMENO-SÁEZ, P.; SENENT-APARICIO, J.; PÉREZ-SÁNCHEZ, J.; PULIDO-VELAZQUEZ, D. A Comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of peninsular Spain. **Water**, v. 10, n. 2, p. 192, 2018. <https://doi.org/10.3390/w10020192>
- KIM, M. et al. Comparative studies of different imputation methods for recovering streamflow observation. **Water**, v. 7, n. 12, p. 6847-6860, 2015. <https://doi.org/10.3390/w7126663>
- KOTHARI, M.; GHARDE, K.D. Application of ANN and fuzzy logic algorithms for stream flow modeling of Savitri catchment. **Journal of Earth System Science**, v. 124, p. 933–943, 2015. <https://doi.org/10.1007/s12040-015-0592-7>
- KOTTEK, M.; GRIESER, J.; BECK, C.; RUDOLF, B.; RUBEL, F. World map of the Köppen-Geiger climate classification updated. **Meteorologische Zeitschrift**, v. 15, n. 3, p. 259-263, 2006.

- KOYCEGIZ, C.; BUYUKYILDIZ, M. Calibration of SWAT and Two Data-Driven Models for a Data-Scarce Mountainous Headwater in Semi-Arid Konya Closed Basin. **Water**, v. 11, n. 1, p. 147, 2019. <https://doi.org/10.3390/w11010147>
- KRAUSE, P.; BOYLE, D. P.; BÄSE, F. Comparison of different efficiency criteria for hydrological model assessment. **Advances in Geosciences**, v. 5, p. 89-97, 2005.
- LIMA, J. E. F. W. Situação e perspectivas sobre as águas do cerrado. **Ciência e cultura**, v. 63, n. 3, p. 27-29, 2011.
- MAKWANA, J. J.; TIWARI, M. K. Intermittent streamflow forecasting and extreme event modelling using wavelet based artificial neural networks. **Water Resources Management**, v. 28, p. 4857-4873, 2014.
- MAKWANA, J. J.; TIWARI, M. K. Hydrological stream flow modelling using soil and water assessment tool (SWAT) and neural networks (NNs) for the Limkheda watershed, Gujarat, India. **Modeling Earth Systems and Environment**, v. 3, n. 2, p. 635-645, 2017.
- MEDRADO, E.; LIMA, J. E. F. W. Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. **Geoderma Regional**, v. 1, p. 59-66, 2014. <https://dx.doi.org/10.1016/j.geodrs.2014.08.003>
- MELLO, C. D.; VIOLA, M. R.; NORTON, L. D.; SILVA, A. M.; WEIMAR, F. A. Development and application of a simple hydrologic model simulation for a Brazilian headwater basin. **Catena**, v. 75, n. 3, p. 235-247, 2008.
- MINNS, A. W.; HALL, M. J. Artificial neural networks as rainfall-runoff models. **Hydrological Sciences Journal**, v. 41, n. 3, 1996. <https://doi.org/10.1080/02626669609491511>
- MONTEIRO, J. A.; STRAUCH, M.; SRINIVASAN, R.; ABBASPOUR, K.; GÜCKER, B. Accuracy of grid precipitation data for Brazil: application in river discharge modelling of the Tocantins catchment. **Hydrological processes**, v. 30, n. 9, p. 1419-1430, 2015.
- MORIASI, D. N.; GITAU, M. W.; PAI, N.; DAGGUPATI, P. Hydrologic and water quality models: Performance measures and evaluation criteria. **Transactions of the ASABE**, v. 58, n. 6, p. 1763-1785, 2015.
- MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853-858, 2000.
- NEITSCH, S. L. et al. **SWAT Model User's Manual**. Texas: Texas A&M University, 2011.
- NÓBREGA, R. L.; GUZHA, A. C.; TORRES, G. N.; KOVACS, K.; LAMPARTER, G.; AMORIM, R. S.; GEROLD, G. Effects of conversion of native cerrado vegetation to pasture on soil hydro-physical properties, evapotranspiration and streamflow on the Amazonian agricultural frontier. **PloS one**, v. 12, n. 6, 2017.
- NOORI, N.; KALIN, L. Coupling SWAT and ANN models for enhanced daily streamflow prediction. **Journal of Hydrology**, v. 533, p. 141-151, 2016. <https://doi.org/10.1016/j.jhydrol.2015.11.050>

- NYEKO, M. Hydrologic Modelling of Data Scarce Basin with SWAT Model: Capabilities and Limitations. **Water Resources Management**, v. 29, p. 81–94, 2015. <https://doi.org/10.1007/s11269-014-0828-3>
- ONS. **Manual de Procedimentos da Operação**. Módulo 10 - Submódulo 10.18. CD-OR.TO.TOC Cadastro de Informações Operacionais Hidráulicas da Bacia do rio Tocantins. 2020. Available at: <http://www.ons.org.br/> Access: 20th Oct. 2020.
- OLIVEIRA, V. A.; DE MELLO, C. R.; BESKOW, S.; VIOLA, M. R.; SRINIVASAN, R. Modeling the effects of climate change on hydrology and sediment load in a headwater basin in the Brazilian Cerrado biome. **Ecological Engineering**, v. 133, p. 20–31, 2019. <https://doi.org/10.1016/j.ecoleng.2019.04.021>
- PAPALASKARIS, T. Artificial Neural Network for Daily Low Stream Flow Rate Prediction of Iokastis Stream, Kavala City, NE Greece, NE Mediterranean Basin. **Environmental Sciences Proceedings, Multidisciplinary Digital Publishing Institute**, v. 2, n. 1, p. 70, 2020. <https://doi.org/10.3390/envirosciproc2020002070>
- PEREIRA, D. R.; MARTINEZ, M. A.; PRUSKI, F. F.; SILVA, D. D. Hydrological simulation in a basin of typical tropical climate and soil using the SWAT model part I: Calibration and validation tests. **Journal of Hydrology: Regional Studies**, v. 7, p. 14–37, 2016. <https://doi.org/10.1016/j.ejrh.2016.05.002>
- PONTES, L. M.; VIOLA, M. R.; SILVA, M. L. N.; BISPO, D. F.; CURI, N. Hydrological modeling of tributaries of cantareira system, Southeast Brazil, with the Swat model. **Engenharia Agrícola**, v. 36, n. 6, p. 1037–1049, 2016. <https://doi.org/10.1590/1809-4430-eng.agric.v36n6p1037-1049/2016>
- PRADHAN, P.; TINGSANCHALI, T.; SHRESTHA, S. Evaluation of Soil and Water Assessment Tool and Artificial Neural Network models for hydrologic simulation in different climatic regions of Asia. **Science of the Total Environment**, v. 20, n. 701, p. 134308, 2020. <http://dx.doi.org/10.1016/j.scitotenv.2019.134308>
- RODRIGUES, J. A.; VIOLA, M. R.; ALVARENGA, L. A.; MELLO, C. R.; CHOU, S. C.; OLIVEIRA, V. A.; UDDAMERI, V.; MORAIS, M. A. Climate change impacts under representative concentration pathway scenarios on streamflow and droughts of basins in the Brazilian Cerrado biome. **International Journal of Climatology**, v. 40, n. 5, p. 2511–2526, 2020. <https://doi.org/10.1002/joc.6347>
- SILVA, J. M. C.; BATES, J. M. Biogeographic patterns and conservation in the south American Cerrado: a tropical savanna hotspot. **Bioscience**, v. 52, n. 3, p. 225–234, 2002.
- SILVA, J. F.; FARIÑAS, M. R.; FELFILI, J. M.; KLINK, C. A. Spatial heterogeneity, land use and conservation in the Cerrado region of Brazil. **Journal of Biogeography**, v. 33, p. 536–548, 2006. <https://doi.org/10.1111/j.1365-2699.2005.01422.x>
- SINGH, A. Modeling Stream Flow with Prediction Uncertainty by Using SWAT Hydrologic and RBNN Models for an Agricultural Watershed in India. **National Academy Science Letters**, v. 39, p. 213–216, 2016. <https://doi.org/10.1007/s40009-016-0436-2>
- TALEBIZADEH, M.; MORIDNEJAD, A. Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. **Expert Systems with applications**, v. 38, n. 4, p. 4126–4135, 2011. <https://doi.org/10.1016/j.eswa.2010.09.075>

- TAN, M. L.; GASSMAN, P.; YANG, X.; HAYWOOD, J. A review of SWAT applications, performance and future needs for simulation of hydro-climatic extremes. **Advances in Water Resources**, v. 143, n. 103662, 2020. <https://doi.org/10.1016/j.advwatres.2020.103662>
- TOCANTINS. Secretaria de Planejamento e da Modernização da Gestão Pública. **Estudo da dinâmica da cobertura e uso da terra do estado do Tocantins**: dinâmica do Tocantins. Palmas, 2012.
- TUKEY, J. W. Comparing individual means in the analysis of variance. **Biometrics**, v. 5, n. 2, p. 99-114, 1949.
- VAN GRIENSVEN, A.; MEIXNER, T.; GRUNWALD, S.; BISHOP, T.; DILUZIO, M.; SRINIVASAN, R. A global sensitivity analysis tool for the parameters of multi-variable catchment models. **Journal of Hydrology**, v. 324, n. 1-4, p. 10-23, 2006. <https://doi.org/10.1016/j.jhydrol.2005.09.008>
- VILANOVA, R. S.; ZANETTI, S. S.; CECILIO, R. A. Assessing combinations of artificial neural networks input/output parameters to better simulate daily streamflow: Case of Brazilian Atlantic Rainforest watersheds. **Computers and Electronics in Agriculture**, v. 167, p. 105080, 2019. <https://doi.org/10.1016/j.compag.2019.105080>
- YASEEN, Z. M.; SULAIMAN, S. O.; DEO, R. C.; CHAU, K.-W. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. **Journal of Hydrology**, v. 569, p. 387-408, 2019. <https://doi.org/10.1016/j.jhydrol.2018.11.069>
- ZAKIZADEH, H.; AHMADI, H.; ZEHTABIAN, G.; MOEINI, A.; MOGHADDAMNIA, A. A novel study of SWAT and ANN models for runoff simulation with application on dataset of metrological stations. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 120, p. 102899, 2020. <https://doi.org/10.1016/j.pce.2020.102899>
- ZHOU, J.; PENG, T.; ZHANG, C.; SUN, N. Data pre-analysis and ensemble of various artificial neural networks for monthly streamflow forecasting. **Water**, v. 10, p. 628, 2018. <https://doi.org/10.3390/w10050628>