Hydrological simulation in a tropical humid basin in the Cerrado biome using the SWAT model

Richarde Marques da Silva, José Carlos Dantas, Joyce de Araújo Beltrão, and Celso A. G. Santos

ABSTRACT

A Soil and Water Assessment Tool (SWAT) model was used to model streamflow in a tropical humid basin in the Cerrado biome, southeastern Brazil. This study was undertaken in the Upper São Francisco River basin, because this basin requires effective management of water resources in drought and high-flow periods. The SWAT model was calibrated for the period of 1978–1998 and validated for 1999–2007. To assess the model calibration and uncertainty, four indices were used: (a) coefficient of determination (R^2); (b) Nash–Sutcliffe efficiency (NS); (c) p-factor, the percentage of data bracketed by the 95% prediction uncertainty (95PPU); and (d) r-factor, the ratio of average thickness of the 95PPU band to the standard deviation of the corresponding measured variable. In this paper, average monthly streamflow from three gauges (Porto das Andorinhas, Pari and Ponte da Taquara) were used. The results indicated that the R^2 values were 0.73, 0.80 and 0.76 and that the NS values were 0.68, 0.79 and 0.73, respectively, during the calibration. The validation also indicated an acceptable performance with $R^2 = 0.80$, 0.76, 0.60 and NS $= 0.61$, 0.64 and 0.58, respectively. This study demonstrates that the SWAT model provides a satisfactory tool to assess basin streamflow and management in Brazil.

Key words | hydrologic modelling, São Francisco basin, SUFI-2 algorithm, SWAT-CUP

Richarde Marques da Silva
José Carlos Dantas
Department of Geosciences,
Federal University of Paraíba,
João Pessoa, PB 58051-900,
Brazil

Joyce de Araújo Beltrão,
Celso A. G. Santos (corresponding author)
Department of Civil and Environmental Engineering,
Federal University of Paraíba,
João Pessoa, PB 58051-900,
Brazil
E-mail: *celso@ct.ufpb.br*

INTRODUCTION

Brazil, along with many other countries, has been severely and repeatedly affected by droughts in terms of extent, intensity and economic impact (Silva *et al.* 2013). Brazil is a country of continental scale and has a broad range of environments (Amazônia, Mata Atlântica, Pantanal, Cerrado (Brazilian savanna), *Caatinga* and Pampa) (IBGE 2016). However, in Brazil few studies for estimating streamflow have been performed on the Cerrado biome, with most attention dedicated to the other biomes, despite the Cerrado being the second most extensive biome in South America (Sano *et al.* 2010), and one of the world's biodiversity hotspots (Grecchi *et al.* 2014). The Cerrado biome covers approximately 204 million ha, which corresponds to 24% of Brazil (Medrado & Lima 2014), and 57% of the Upper São Francisco River basin. Since the 1970s, the Cerrado biome has suffered loss of vegetation due to agriculture and pasture expansion (Silva *et al.* 2006), although understanding of these land cover changes and their impacts on streamflow in the Cerrado biome is limited (Beuchle *et al.* 2015). In contrast, the Amazônia, Mata Atlântica and Pantanal biomes have been the focus of significant monitoring since the 1980s. Only recently have some studies focused on assessing hydrologic modelling in the southeastern region. These have been in small basins (Durães *et al.* 2011; Pinto *et al.* 2013; Pereira *et al.* 2014, Pereira 2016a), which do not represent streamflows at the biome-scale and,

doi: 10.2166/nh.2018.222

particularly, impact on inflows to the Três Marias reservoir. Improving the knowledge base on hydrological processes in the Cerrado biome is therefore important for the national water resources management.

Mapping of land-cover change and hydrologic modelling are important tools for water resource planning and management in river basins in a changing climate (Beskow et al. 2009), particularly in basins with large rainfall variability and different land uses, such as the Upper São Francisco River basin. This basin of 48,679 km^2 drains the main watercourse of Minas Gerais State and is located in the southeastern region of Brazil, characterized by a tropical humid climate. The Upper São Francisco River basin is a strategic area in terms of water resources and includes the Três Marias reservoir, with a capacity of 20 billion m^3 that plays an important role in regulating river discharge, and is used for multiple purposes: (a) hydropower generation; (b) navigation; and (c) water supply for population, industries and irrigation projects (Fan et al. 2014). Furthermore, this river basin is of the utmost importance for regional development, supporting diverse agricultural and industrial enterprises (Versiani et al. 2009). The basin has a particularity: a non-intensive, but constant, cattle farming, which occurs largely in the Cerrado biome, in addition to the intensive agriculture, which influences the surface soil characteristics and consequently influences the hydrological regime of the basin. There are approximately 300,000 ha of irrigated land in the region and fruit production for export (grapes and mangoes, among others), dry-land farming, cattle-raising and mining are among the main economic activities (CEMIG 2015). This land use, together with intensive agriculture, impacts surface soil characteristics and influences the hydrological regime of the basin.

Recently the southeastern region of Brazil has experienced significant rainfall deficits, and consequently, recurring droughts in the Upper São Francisco River basin (Coelho et al. 2016), e.g., 1998, 1999, 2001 and 2012. In these years, the region suffered from severe water shortages leading to depleted water levels in reservoirs such as Três Marias, which had, in turn, a direct impact on hydropower energy generation and water availability for public consumption and agriculture. Water scarcity refers to the relative water shortage of the supply system, which can be caused by climate or by human actions due to demand exceeding supply – something that can also be linked to population growth and higher living standards (Chen & Li 2016). Quantification of current and future water availability is an important aspect of water management in order to ensure food security, human health, environmental health and economic development.

Modelling techniques have been developed to quantify water scarcity and availability (Chen & Li 2016), and hydrological models can be used for simulating past, present and future land-use change impact at basin scales based on different scenarios (Li et al. 2015). The Soil and Water Assessment Tool (SWAT) model has been used in a wide range of studies across Brazil (Durães et al. 2011), but there has been limited testing in Brazilian savannah conditions. Lima et al. (2014) also note that further development of the SWAT soil parameter database to support a greater range of Brazilian physical characteristics was needed, and especially for the Cerrado. Therefore, the aims of this paper are to firstly develop a consistent soil parameter database and secondly to assess performance of a model in simulating streamflow in the Upper São Francisco River basin of the Cerrado biome, using the SWAT model.

MATERIALS AND METHODS

Study area

The research was conducted in the Upper São Francisco River basin (48,679 km^2), located in the central region of Minas Gerais state, in southeastern Brazil, between latitudes 21°10′S and 18°10′S, and longitudes 46°40′W and 43°40′W (Figure 1). Local climate types according to the Köppen classification (Alvares et al. 2013) include Cwa (with dry winter and hot summer), Cwb (with dry winter and temperature summer) and Aw (tropical zone with dry winter), with a well defined rainy season between December and March and average annual rainfall ranges from 800 to 1,200 mm (Rao et al. 2016).

The local biome is Cerrado, which is a seasonal ecosystem characterized by distinct dry and wet seasons. The typical vegetation of this biome is tall and dense evergreen gallery forests, ranging from closed to open canopy deciduous and semi-deciduous forest with a maximum height of

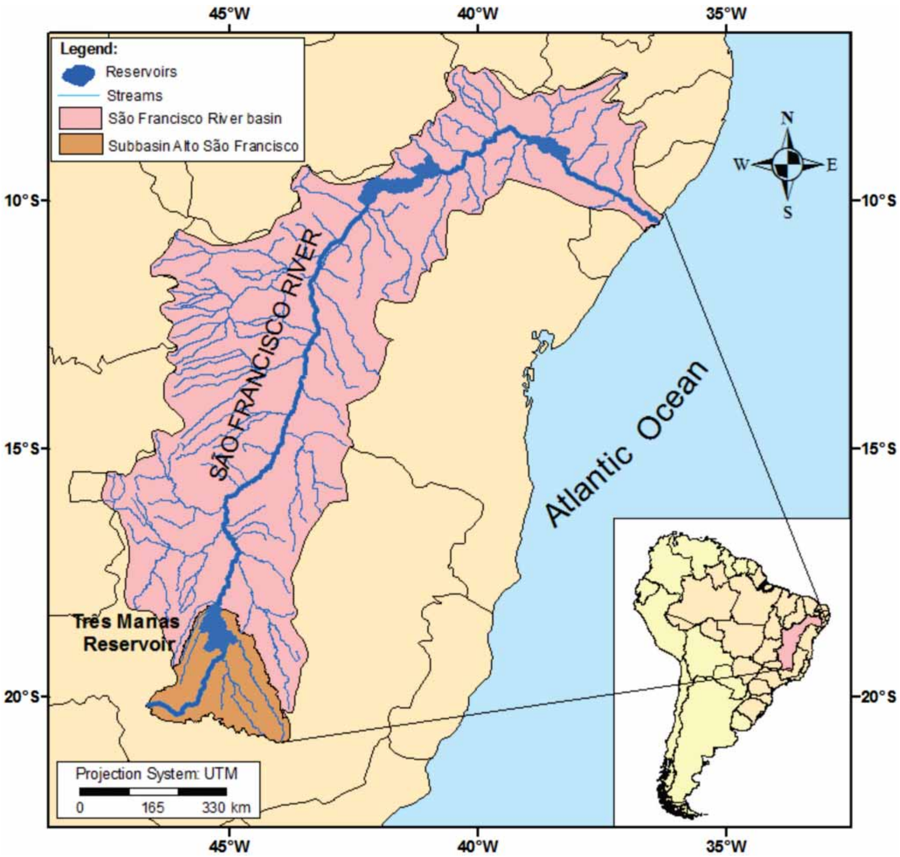

Figure 1 | Geographical location of the São Francisco River basin and Upper São Francisco sub-basin, Brazil.

15 m (Klink & Machado 2005). The main land uses in the Upper São Francisco River basin have been classified as cerrado and agriculture.

Datasets and input data

The rainfall database was made up of historical daily rainfall series from 12 rainfall stations. Three streamflow stations and three weather stations were also used (Table 1 and Figure 2). Rainfall and streamflow data were obtained from the National Water Resources Information System (SNIRH) (available at http://www.ana.gov.br/portalsnirh). Observed meteorological data (daily values of total rainfall, mean wind speed, maximum and minimum temperatures, mean relative humidity and solar radiation) measured from 2 m height from the three weather stations (Table 1 and Figure 2) were obtained from the Instituto Nacional de Meteorologia (INMET) (available at http://www.inmet.gov.br).

Figure 3(a) shows the land use and cover map derived from satellite imagery, captured by the TM/Landsat 5 satellite (orbit/point 218/74), with 30×30 m spatial resolution, acquired on September 21, 2003, accessible at the Instituto Nacional de Pesquisas Espaciais – INPE (available at: http://www.inpe.br). The images were classified using the K-means unsupervised digital image classification method (Santana *et al.* 2014), with field verification to identify and confirm land use. The identified land uses were: (a) cerrado; (b) agriculture; (c) pasture; (d) bare soil; (e) urban area; and (f) water bodies. Reclassification of the land-use map was undertaken and the respective model parameters were selected from the SWAT database corresponding to the specific land-use and cover (LUC) types. The areas of each LUC classification system are shown in Table 2.

Table 1 | Geographic characteristics of the rainfall, streamflow and weather stations used in the study

Stations	Type	Latitude (degree)	Longitude (degree)	Elevation (m)	Period of data
Araxá	Rainfall	−19.60	−46.94	1,023	1978–2007
Bambuí	Rainfall	−20.03	−45.00	661	1978–2007
Barbacena	Rainfall	−21.25	−43.76	1,126	1978–2007
Belo Horizonte	Rainfall	−19.93	−43.93	915	1978–2007
Florestal	Rainfall	−19.88	−44.41	760	1978–2007
Ibirité	Rainfall	−20.01	−44.05	814	1978–2007
João Pinheiro	Rainfall	−17.73	−46.17	760	1978–2007
Patos de Minas	Rainfall	−18.51	−46.43	940	1978–2007
Pirapora	Rainfall	−17.35	−44.91	505	1978–2007
Pompeu	Rainfall	−19.21	−45.00	690	1978–2007
São Sebastião do Paraíso	Rainfall	−20.91	−47.11	820	1978–2007
Sete Lagoas	Rainfall	−19.46	−44.25	732	1978–2007
Porto das Andorinhas	Streamflow	−19.28	−45.29	532	1978–2007
Pari	Streamflow	−20.18	−44.89	710	1978–2007
Ponte da Taquara	Streamflow	−19.42	−44.55	627	1978–2007
Bambuí	Weather	−20.03	−45.00	661	1978–2007
Araxá	Weather	−19.60	−46.94	1,023	1978–2007
São Sebastião do Paraíso	Weather	−20.91	−47.11	820	1978–2007

Figure 3(b) presents a 30 m resolution digital elevation model (DEM) of the Upper São Francisco River basin, obtained from the INPE (available at: http://www.topodata.inpe.br). The EMBRAPA soil map at a scale of 1:250,000 (Figure 3(c)) (available at: http://mapoteca.cnps.embrapa.br) was applied and the soil data properties obtained from the EMBRAPA soil properties database (EMBRAPA 2015) were used. The soil properties data (soil texture, available water content, hydraulic conductivity, bulk density and organic carbon content) for different layers of each soil type were collected and inserted in a database of SWAT model. The major soil groups in the basin are Cambisol (43%) and Red Latosol (23%).

From the natural river network, the topography of the basin and the distribution of rainfall stations, the study area was divided into 22 sub-basins and 180 HRUs (Figure 3(d)) for the purposes of routing water and sediment (Santos *et al.* 2011). SWAT divides a basin into sub-basins, each of which is linked through a stream channel and further discretized into hydrologic response units (HRUs), that enables the model to reflect the hydrologic conditions for different slope, land cover and soil types (Patel & Srivastava 2013). A HRU is a unique combination of soil and vegetation type in a sub-basin, and the model simulates hydrological processes at the HRU level and aggregates these results to the catchment scale by applying a weighted average to the HRU results (Devkota & Gyawali 2015). In defining the HRUs, the minor land use and cover, soil types and slope were ignored by setting a threshold of 15% to avoid an unnecessarily large number of HRUs (Neitsch *et al.* 2005). The drainage areas accounting for each one of the three runoff gauges used in the simulation are: (a) Porto Andorinhas = 13,005 km^2; (b) Pari = 4,521 km^2; and (c) Ponte da Taquara = 7,179 km^2.

Application of the SWAT model and automatic calibration

The SWAT model is a semi-distributed, time-continuous basin simulator on a daily, monthly and annual time step, for hydrological modelling (Arnold *et al.* 1998). The hydrologic processes considered in the model comprise

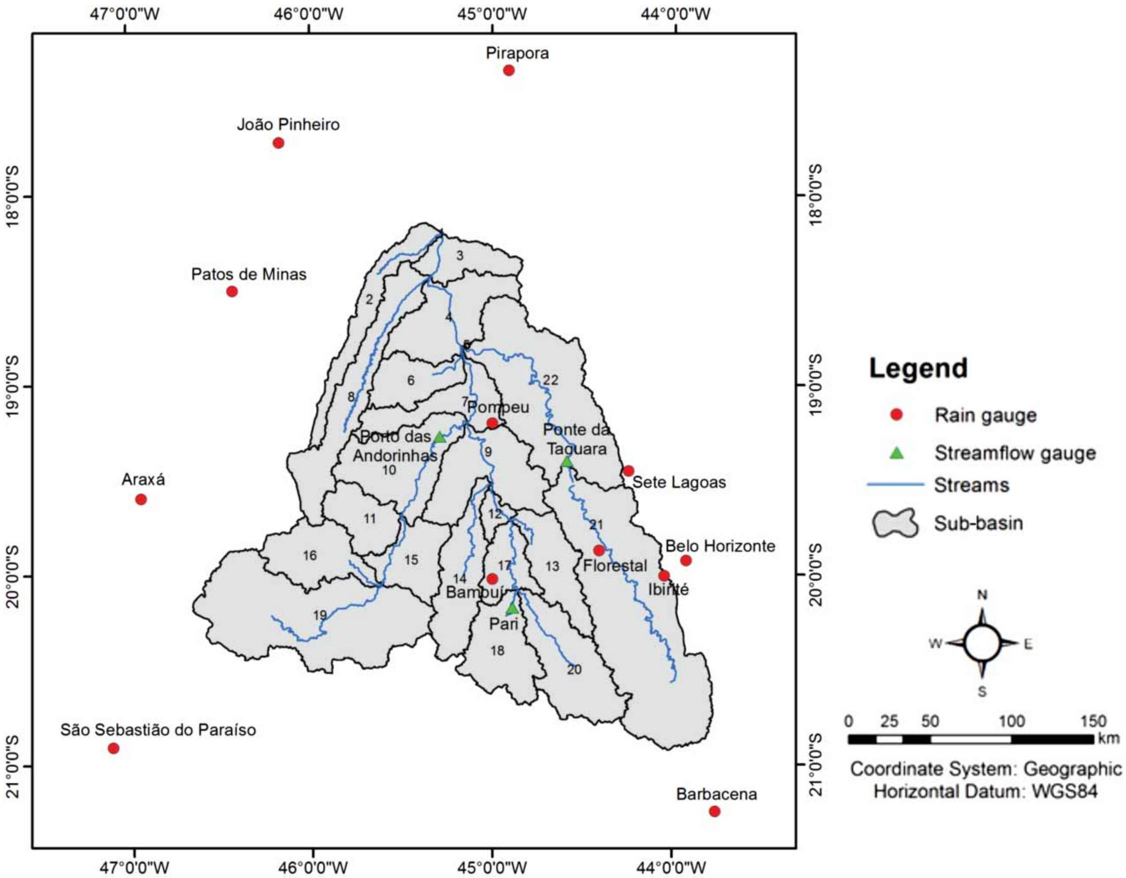

Figure 2 | Upper São Francisco River basin with its sub-basins, and rainfall and streamflow stations.

evaporation, percolation, infiltration, groundwater flows and lateral flows (Neitsch *et al.* 2005). The modified Soil Conservation Service (SCS) curve number (CN) method is used to estimate the surface runoff volume (Narsimlu *et al.* 2015). The water balance, Equation (1), which rules the hydrological cycle components of the SWAT is (Neitsch *et al.* 2005):

$$SW_t = SW_0 + \sum_{i=1}^{t} \left(R_i - Q_i - ET_i - P_i - Q_{g_i} \right) \quad (1)$$

where SW_t is the final soil water content (mm), SW_0 is the initial soil water content on day i (mm), t is the time (days), R_i is the amount of rainfall on day i (mm), ET_i is the amount of evapotranspiration on day i (mm), Q_{gi} is the amount of return flow on day i (mm), Q_i is the amount of daily surface runoff (mm), and P_i is the amount of water entering the vadose zone from the soil profile on day i (mm). Figure 4 shows the schematic framework of the SWAT with input data (maps and databases), discretization process and hydrological processes output.

The SWAT model has more than 40 calibration parameters and effective calibration begins by developing a proper mechanism for reducing the number of parameters to be calibrated (Xu *et al.* 2009). Parameter values that were not derived from measurements or the literature were assigned based on either automated or manual calibration. In the initial phase, parameter sensitivity was evaluated combining the methods One-factor-At-a-Time (OAT) and Latin Hypercube (LH) (van Griensven *et al.* 2006), identifying and classifying the parameters that have the most significant impact on model outputs (Panagopoulos *et al.* 2011; Wu & Chen 2015). Subsequently, an automatic calibration based on the sensitivity analysis of model parameters called SWAT-CUP (SWAT Calibration Uncertainty Programs) (Abbaspour *et al.* 2004), was used. In the SWAT-CUP, the SUFI-2 algorithm (Abbaspour 2007)

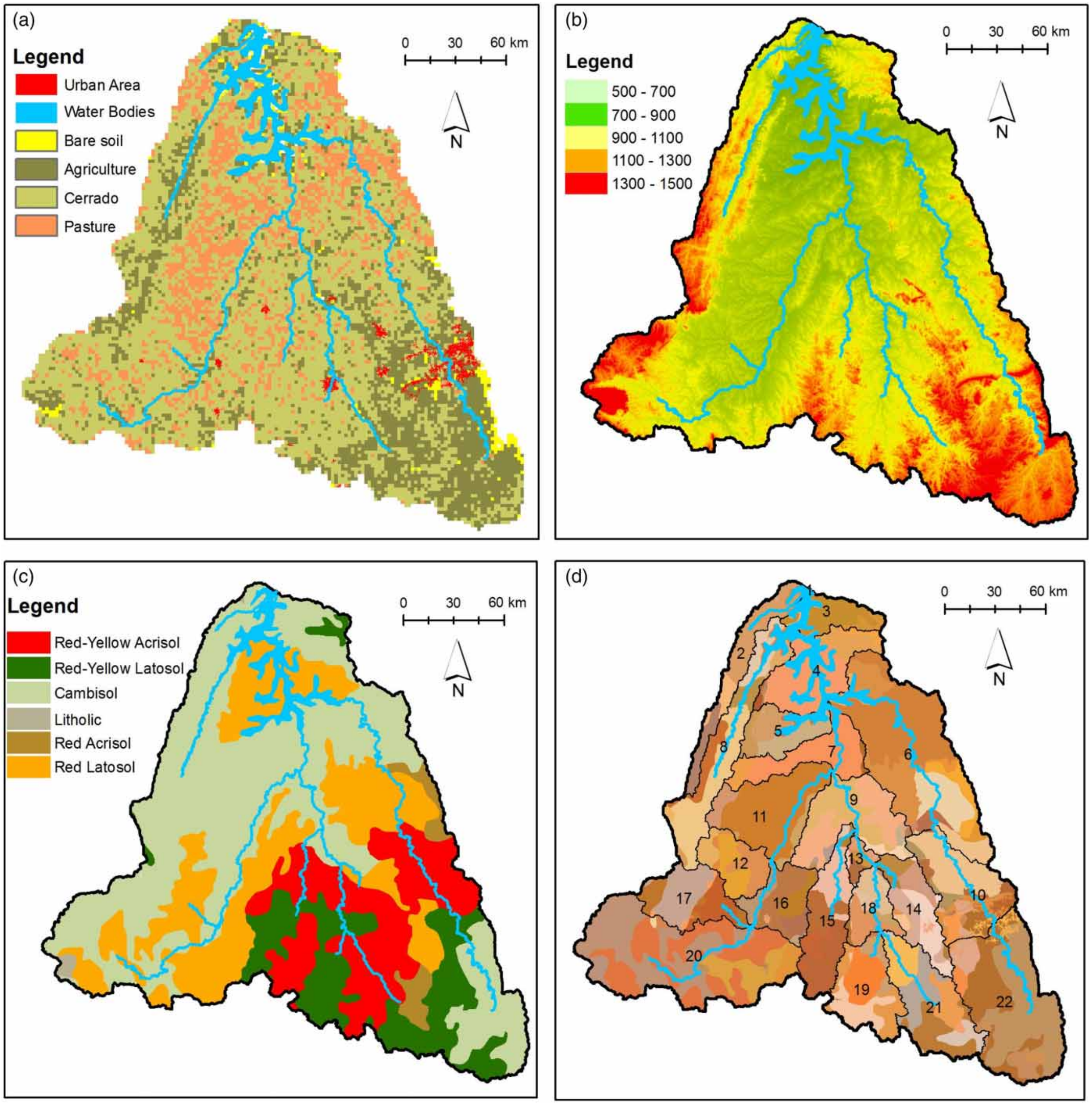

Figure 3 | (a) LUC map, (b) digital elevation model, (c) soil map, and (d) hydrologic response units (HRUs), of the study area.

is used to investigate the sensitivity and uncertainty in various SWAT parameters related to streamflow. SUFI-2 is a widely used algorithm that estimates the sensitivity and uncertainty of a hydrological model, and is powerful in communicating the reliability of model predictions (Narsimlu *et al.* 2015). Latin Hypercube sampling is carried out and the corresponding objective function was used to calculate the responsive parameter sensitivities (Yang *et al.* 2008). In order to estimate the level of significance of each parameter, the *t*-test and the *p*-values were used to provide a measure and the significance of the sensitivity, respectively.

For automatic calibration and uncertainty analysis, a total of 19 SWAT model parameters with high calibration sensitivity for streamflow prediction were adjusted. The calibrated parameters and the ranges used are shown in Table 3, ranked from the most to least sensitive parameters. Once the

Table 2 | LUC and soil classes for the Upper São Francisco River basin

SWAT class	Description	Type	Area (km^2)	% of Area
RNGB	Cerrado	LUC	26,128	53.67
AGRL	Agriculture	LUC	11,452	23.53
PAST	Pasture	LUC	7 848	16.12
WATR	Water	LUC	1 519	3.12
BARR	Bare soil	LUC	1 167	2.40
URBN	Urban area	LUC	565	1.16
Total	–	–	**48,679**	**100.00**
CAMB	Cambisol	Soil	21,046	43.23
RLAT	Red Latosol	Soil	11,327	23.27
RYAC	Red-yellow Acrisol	Soil	7 523	15.45
RYLA	Red-yellow Latosol	Soil	6 099	12.53
WATR	Water		1 519	3.12
RACR	Red Acrisol	Soil	1 009	2.07
LITH	Litholic	Soil	156	0.32
Total	–	–	**48,679**	**100.00**

model parameters were optimized for calibration, model validation was performed based on monthly streamflow data.

Calibration and validation

The SWAT model was applied for three flow stations (Porto das Andorinhas, Pari and Ponte da Taquara). Calibration and validation of SWAT, facilitated by the sensitivity analysis, were performed for streamflow on a monthly basis. Model calibration was performed for the period of January 1978 to December 1998 (i.e. 20 years of data for calibration), with three years of warm up (1975–1977). SWAT was validated for the Upper São Francisco by carrying out a Split Sample Test, where the model was calibrated for the period 1999 to 2006, and the optimized parameters were applied to the period of January 1999 to December 2007 (validation).

For temporal bias, four indicators were used: (a) coefficient of determination (R^2); (b) Nash–Sutcliffe efficiency (NS); (c) p-factor; and (d) r-factor. The p-factor, which is the percentage of observations bracketed by the 95% prediction uncertainty (95PPU), was used to calculate the uncertainties associated with the SWAT model calibration (Narsimlu *et al.* 2015), and the r-factor (average width of the 95PPU band divided by the standard deviation of the

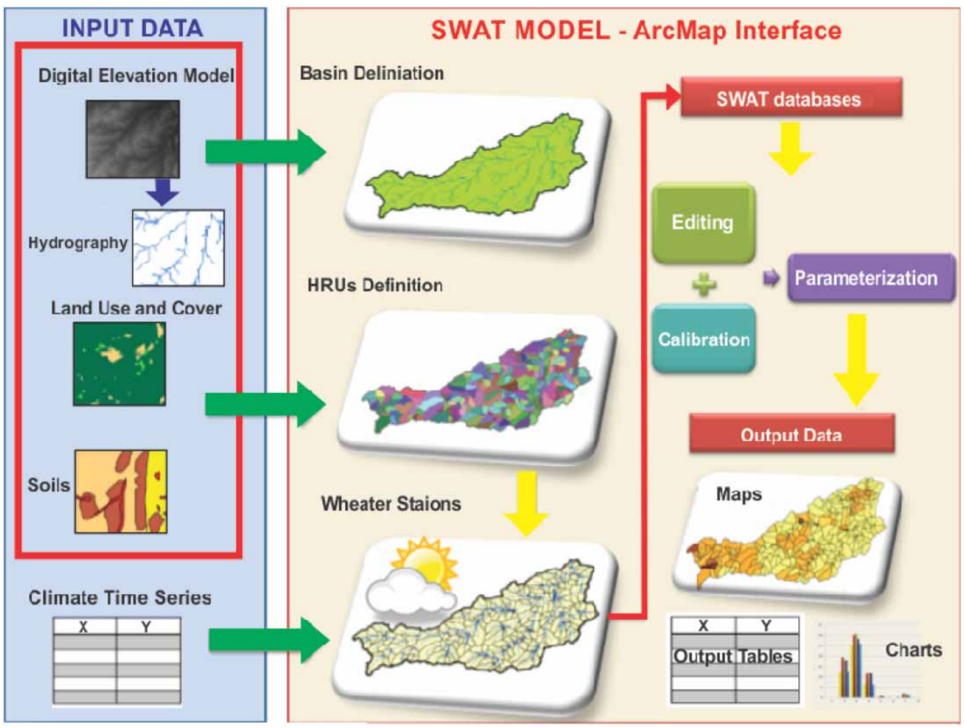

Figure 4 | Schematic framework of the SWAT model.

Table 3 | SWAT parameters used for calibration and the higher ranks of the parameters

Parameters	Parameter definition	Ranks of the parameters
GW_REVAP	Revap coefficient	1
GW_DELAY	Groundwater delay time	2
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur	3
SOL_AWC	Soil available water storage capacity	4
SOL_Z	Depth from soil surface to bottom of layer	5
CN2	Curve number for moisture condition II	6
ALPHA_BF	Base flow alpha factor	7
ESCO	Soil evaporation compensation factor	8
CH_N2	Manning's *n* value for the main channel	9
CH_K2	Main channel conductivity	10
SLSUBBSN	Average slope length	11
CANMX	Maximum canopy storage	12
SOL_ALB	Moist soil albedo	13
SOL_K	Soil hydraulic conductivity	14
RCHRG_DP	Deep aquifer percolation fraction	15
SURLAG	Surface runoff lag coefficient	16
BIOMIX	Biological mixing efficiency	17
EPCO	Plant uptake compensation factor	18
REVAPMN	Threshold water in shallow aquifer	19

measured variable) were used for evaluation of model performance. The r-factor is given by:

$$\text{r-factor} = \frac{(1/2)\sum_{t_i}^{M}\left(y^{M}_{t_i,97.5\%} - y^{M}_{t_i,2.5\%}\right)}{\sigma_{obs}} \quad (2)$$

where $y^{M}_{t_i,97.5\%}$ and $y^{M}_{t_i,2.5\%}$ are the upper and lower boundaries of the 95PPU; σ_{obs} is the standard deviation of the observed data.

The Latin hypercube sampling method was employed for 95PPU and for obtaining the final cumulative distribution of the model outputs. These are calculated at the level of 2.5 and 97.5% prediction limit. At the start of calibration, SUFI-2 assumes a large parameter uncertainty and then decreases this uncertainty through the p-factor and the r-factor performance statistics. The range of the p-factor varies from 0 to 1, with values close to 1 indicating a very high model performance and efficiency, while the r-factor is the average width of the 95PPU band divided by the standard deviation of the measured variable and varies in the range 0–1 (Yang *et al.* 2008).

The R^2 describes the degree of collinearity between simulated and measured data, and defines the proportion of the variance in measured data explained by the model. The NS is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance. This index indicates how close observed values are to simulated values. Nash-Sutcliffe model efficiency (Nash & Sutcliffe 1970) and R^2 can be calculated using the following equations:

$$NS = 1 - \left(\sum_{i=1}^{n}(Q_o - Q_s)^2 \Big/ \sum_{i=1}^{n}(Q_o - \bar{Q}_o)^2\right) \quad (3)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(Q_o - \bar{Q}_o)(Q_s - \bar{Q}_s)}{\sqrt{\sum_{i=1}^{n}(Q_o - \bar{Q}_o)^2\sum_{i=1}^{n}(Q_s - \bar{Q}_s)^2}}\right)^2 \quad (4)$$

where Q_o and Q_s are the measured and simulated data, respectively, $\bar{Q}_o$ and $\bar{Q}_s$ are the mean measured and simulated data, respectively, and n is the total number of data records.

RESULTS AND DISCUSSION

Sensitivity analysis

Table 4 presents details of the parameters used for the sensitivity analysis and the most sensitive parameters. The results of the sensitivity analysis confirmed that all 19 sensitive parameters are applicable to surface runoff, groundwater, channel routing, and soil properties. Several SWAT parameters were set during the model calibration: SLSUBBSN, GWQMN, CN2, CH_K2, CH_N2, ALPHA_BF, ESCO, EPCO, REVAPMN, SOL_AWC, SOL_K, SOL_BD, GW_DELAY, GW_REVAP, and HRU_SLP. The larger absolute values are more sensitive than the lower ones, while a value closer to zero has more significance. In this study, the result of the global sensitivity analysis with the *t*-test suggests that the most sensitive parameters are GW_REVAP, GW_DELAY, GWQMN, SOL_AWC, SOL_Z, CN2, ALPHA_BF and ESCO. The results show that the most sensitive parameters are those representing the groundwater, base flow, surface runoff and soil properties, linked to basin rainfall seasonality, and diversity of lithology, landscapes, soil, and land use, all of which influence soil available water storage capacity.

SWAT calibration and validation

Figure 5 compares the measured and simulated monthly runoff values at the Porto das Andorinhas station for both the calibration and validation periods. The simulated and measured monthly streamflows at the Porto das Andorinhas stations are similar during most of the calibration period.

Table 4 | Sensitive SWAT model parameters included in the final calibration, with their variations (initial and final), *p*-values and t-stat for the studied gauges

			Range		Fitted values			
Parameters	**t-stat**	***p*-value**	**Min. value**	**Max. value**	**Porto das Andorinhas**	**Pari**	**Ponte de Taquara**	**Method**
GW_REVAP	−46.39	0	0.02	0.2	0.18614	0.03926	0.15698	Replace
GW_DELAY	−11.59	0	−30	60	45.33	−5.25	−1.83	Added
GWQMN	−10.40	0	0	1000	415	165	839	Replace
SOL_AWC	−8.633	0	−0.25	0.25	−0.0195	−0.1625	0.1585	Multiply
SOL_Z	−8.065	0	−0.25	0.25	0.1565	0.1865	0.1585	Multiply
CN2	6.283	0	−0.1	0.1	0.0354	−0.0978	−0.0106	Multiply
ALPHA_BF	7.335	0	0	1	0.023	0.671	0.005	Replace
ESCO	23.34	0	0.5	1	0.5435	0.9985	0.6865	Replace
CH_N2	−2.837	0.004	0	0.3	0.1743	0.0051	0.2481	Replace
CH_K2	−2.724	0.006	0	5	1.195	4.415	1.965	Replace
SLSUBBSN	−1.299	0.194	−0.25	0.25	0.2045	0.0755	−0.0665	Multiply
CANMX	1.261	0.207	2	5	4.241	2.915	2.453	Replace
SOL_ALB	−0.919	0.358	−0.25	0.25	0.0705	0.2135	0.1955	Multiply
SOL_K	−0.897	0.37	−0.25	0.25	0.2085	−0.2465	0.0575	Multiply
RCHRG_DP	0.734	0.463	−0.04	0.05	0.03929	0.02327	−0.01669	Multiply
SURLAG	0.732	0.464	0	24	1.896	11.304	16.632	Replace
BIOMIX	−0.659	0.509	0	1	0.893	0.139	0.573	Replace
EPCO	0.431	0.666	0	1	0.577	0.839	0.229	Replace
REVAPMN	−0.221	0.824	0	10	2.11	2.65	5.51	Replace

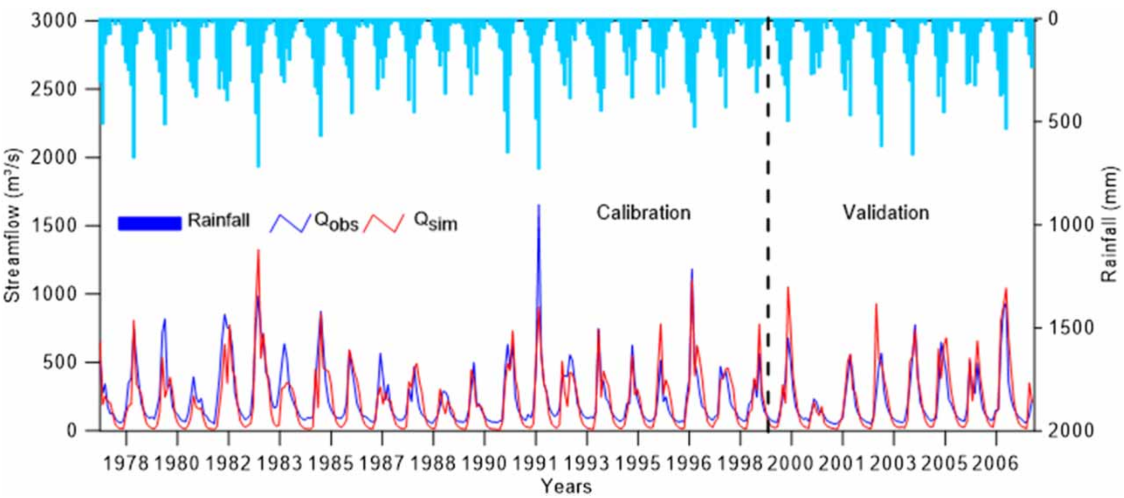

Figure 5 | Simulated and observed monthly runoff and observed hyetograph at the Porto das Andorinhas station.

The simulated streamflow follows a similar trend for both periods. Monthly simulation results are satisfactory and the model was able to track the seasonal trends very well. The results reflect the seasonal change and match the daily streamflow to a reasonable degree. However, the results indicate intermittent over-/underestimation. Most peaks in the summer (September, October and November) were significantly underestimated, while the mean monthly streamflow in March to May was obviously overestimated in most cases. The values of the streamflow reveal the greatest variation of magnitude in the monitored and simulated events, clearly indicating an imbalance in the infiltration process and formation of streamflow.

The hydrographs (Figure 5) in the validation period also highlighted this typical phenomenon. Flood peaks in summer were greatly overestimated. The largest differences were observed for months in which high rainfall totals were registered. In these cases, the model tended to overestimate the runoff and, as a consequence, to increase the simulated discharge peaks.

Figure 5 clearly shows that the model satisfactorily simulates the monthly runoff and that the hydrographs are in reasonable agreement with the rainfall pattern. The values of the Nash–Sutcliffe efficiency (NS equal to 0.68 for calibration and 0.61 for validation) and the coefficient of determination (R^2 equal to 0.73 for calibration and 0.80 for validation) and the small volume difference between the observed and simulated values of 9% (87.64 $m^3 s^{-1}$ for observed monthly runoff and 79.74 $m^3 s^{-1}$ for simulated monthly runoff) show a good predictive capability of the model (according to Moriasi *et al.* 2007) at the Porto das Andorinhas station. Other studies using the SWAT model related NS values ranged from 0.3 to 0.9, depending on the drainage area of basin, the time interval of the simulation and the available database (Fukunaga *et al.* 2015).

Overall, the streamflow simulated by the SWAT model fits the observed well, albeit with some difficulties in simulating some streamflow peaks both in the calibration and validation periods (Figure 5). This same problem is related to the representation of the temporal runoff and was found by many authors in basins near to Upper São Francisco (Durães *et al.* 2011; Pinto *et al.* 2013; Pereira *et al.* 2014, 2016a).

Figure 6 shows the comparison between the observed and simulated streamflow for the calibration and validation periods at the Pari station. This figure shows that the simulated values are again in good agreement with the observed values for calibration and validation periods. This finding is justified by the high NS of 0.79 and R^2 of 0.81 and the small volume difference of 8.6%. However, for the validation, the model overestimates the runoff ($NS = 0.64$ and $R^2 = 0.76$). The results with the SWAT model showed that the baseflow component in the basin was not modeled well. This low model predictability is also due to: (a) changes in the land use, and climate variability in the basin, which were not able to be incorporated into

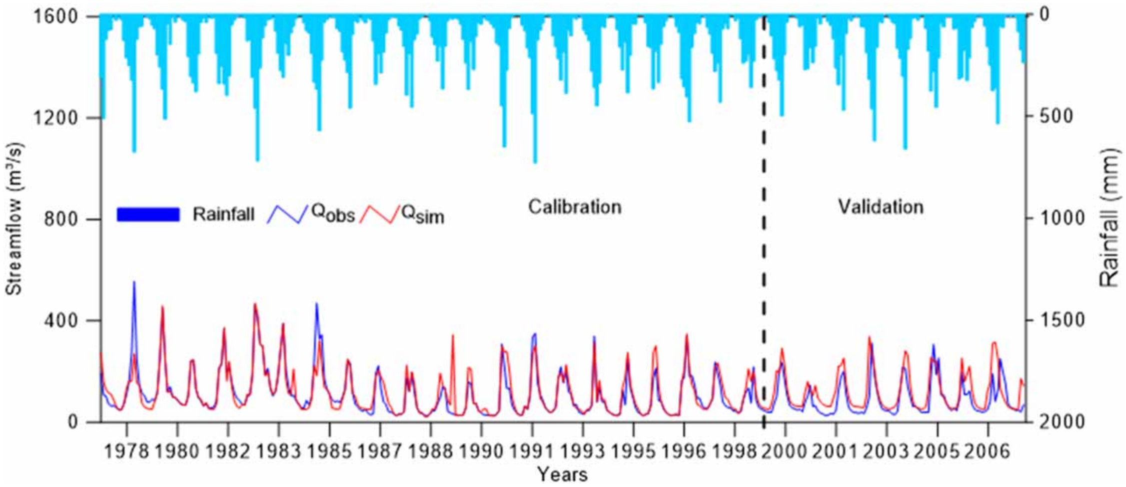

Figure 6 | Simulated and observed monthly runoff and observed hyetograph at the Pari station.

the hydrologic simulations, as also shown by Rodrigues *et al.* (2015); (b) changes in the soil water infiltration and retention patterns; and (c) increase in extreme events which caused a volume reduction in the streamflow time series (Pruski *et al.* 2004).

The results show that the model had a base flow underestimation bias during the calibration period. Combined with the temporal rainfall variation, which is difficult to represent through a daily time step, this made the response of the SWAT model in simulating some peak streamflow difficult. Nevertheless, this is considered a reasonable monthly model performance, especially when compared with the values cited in the studies mentioned earlier.

Figure 7 presents the hydrograph of the calibration and validation periods for the Ponte da Taquara station. A close relationship between the simulated and observed monthly streamflow indicates a good performance of the model (Moriasi *et al.* 2007), with R^2 values of 0.76 and 0.60, and NS values of 0.73 and 0.58 for calibration and validation, respectively. Again, the monthly simulation results are satisfactory and the model was also able to track the seasonal trends very well. Although the results reveal that the model satisfactorily represented the simulated flow, underestimations are noted in the recession limb of the hydrograph, possibly related to the high spatial variability of climatic data produced with the weather generator to replace missing

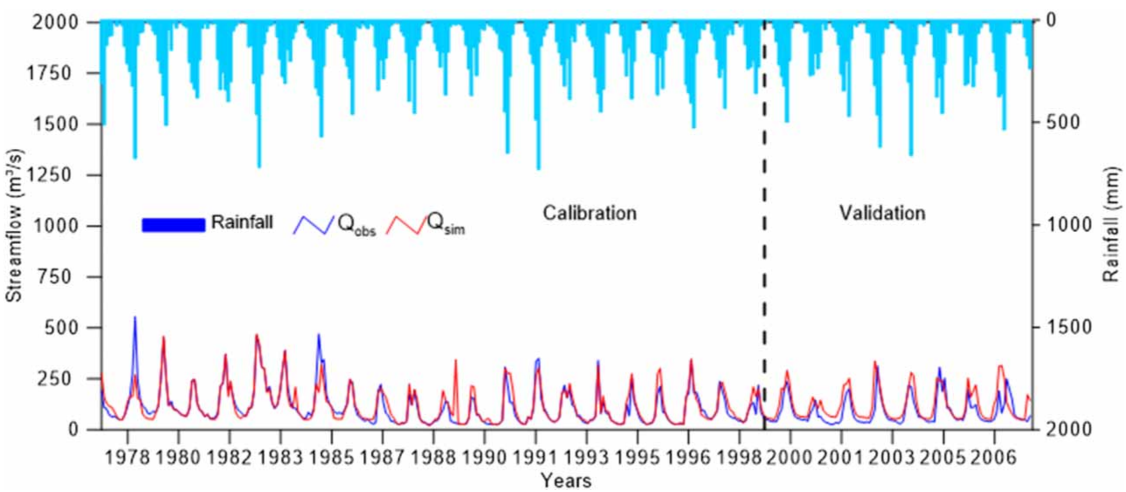

Figure 7 | Simulated and observed monthly streamflow and observed hyetograph at the Ponte da Taquara station.

values and also due to the influence of the aquifer systems, which are necessary for the model application. Such marked differences in the case of our study could be due to the location and low number of hydrometeorological stations used in relation to the size of the study area, as they would not have been able to record some local high-intensity rainfall events. Alternatively this could be the result of intra-annual variability of rainfall during the calibration years. Mello *et al.* (2008), applying the same model for sub-basins in Alto Rio Grande basin, found NS values above 0.7 in the calibration and validation phases. Rostamian *et al.* (2008) reported that a SWAT model is not designed to simulate such an extreme event and the model usually under predicts the largest flow events increasing the uncertainty concerning the calibration of extreme events.

The modeled hydrograph showed goodness of fit to observations, which means there was good estimation of discharge in that phase, observing model sensitivity to peak discharges. Thus, the model succeeded in capturing oscillations brought about by rainfall. The model slightly underestimated some low flows and the agreement between observed and simulated flows was stronger in the monthly time scale. The model also performed well for the validation period, and these results showed that the hydrological processes were reasonably represented, suggesting that this model could be used for water impact assessment, as also obtained for a daily scale by Durães *et al.* (2011) for Paraopeba basin (sub-basin of Upper São Francisco) and Pereira *et al.* (2016b) for Pomba River Basin, which are located in southeastern region of Brazil.

Rainfall distribution in SWAT is undertaken by associating each sub-basin with the nearest rainfall gauging station closest to its centroid, and because the sub-basin has only four rainfall gauging stations upstream from the control section in Ponte da Taquara, actual spatial distribution of rainfall may not have been representative.

Reasonably high values of NS and R^2 for monthly streamflow in the Porto das Andorinhas and Pari stations indicate better prediction for both stations during the calibration period. Although an overestimation of monthly streamflow by the model was beyond the satisfactory level of acceptance, the model can be considered good considering the overall statistics. In general, the model accurately tracked the observed streamflow for the time periods, although some monthly runoff peaks were over-predicted during calibration and validation, with smaller over-prediction the during validation.

Uncertainty in streamflow prediction

We used the default lower and upper bound parameter values, which were taken directly from the SWAT users' manual (Neitsch *et al.* 2005). Uncertainty analysis results of SUFI-2 during the calibration and validation periods at the three flow stations are shown in Figure 8. In these figures, the shaded region (95PPU) contains all uncertainties from the different sources.

The results for Pari flow station (Figure 8(b)) show that 84% of the observed data were bracketed by the 95PPU. Most of the data were captured by the 95PPU. The SWAT model was capable of simulating large flows and extreme events in the Upper São Francisco River basin, except for Ponte da Taquara flow station. Some peak values were missing from the 95PPU band. This shows that SUFI-2 did not capture the observed data well for Ponte da Taquara and Porto das Andorinhas stations and had high uncertainty for simulated peak values, which shows that calibration of the SWAT model in this basin is challenging due to the uncertainties that are driven by the runoff process, which are not totally understood.

The results indicated that the p-factor values were 0.89, 0.81 and 0.79 and that the r-factor values were 1.53, 2.34 and 2.95, respectively, during the calibration. The validation also indicated an acceptable performance with p-factor = 0.87, 0.86, 0.81 and r-factor = 1.72, 2.35, 3.44, respectively. In all stations, the model mostly shows large uncertainties at extreme events during the calibration and validation periods. These results were generally better for the Pari station in terms of the percentage of data being bracketed (p-factor), but the uncertainties are larger as expressed by the r-factor in all stations. This has resulted in large uncertainty in discharge peaks for all stations. Although the simulation of monthly runoff for the Porto das Andorinhas station was satisfactory during the calibration period, the SWAT model exhibited large uncertainties for the validation period.

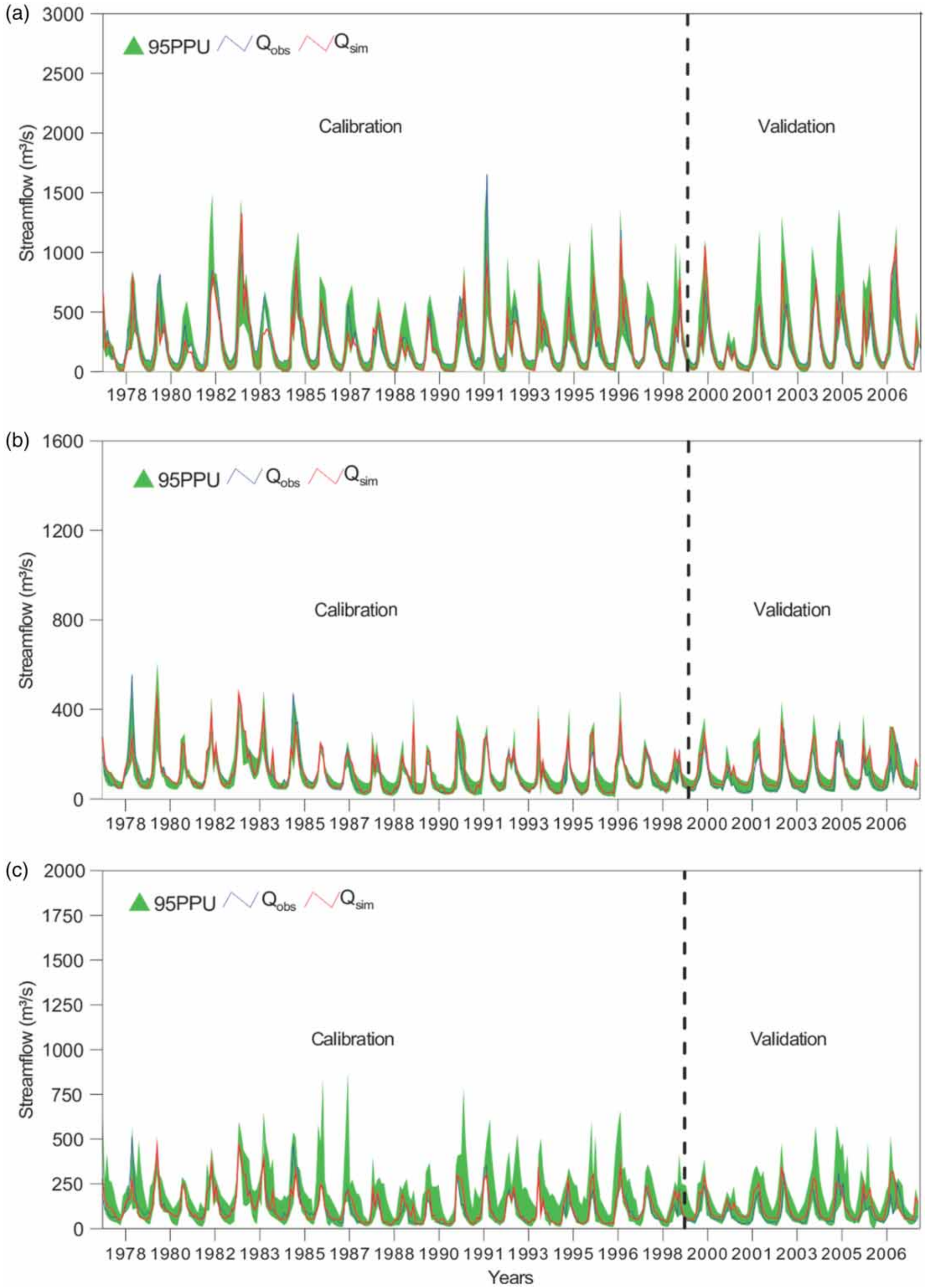

Figure 8 | 95% probability uncertainty plot observed and simulated streamflow for stations: (a) Porto das Andorinhas, (b) Pari and (c) Ponte da Taquara.

The parameter uncertainties were tolerable when the parameter ranges of the NS and R^2 reached the desired limits. When the NS value is >0.60, the results are satisfactory, and once NS is >0.75, the simulation results are good (Nash & Sutcliffe 1970). For the results during the calibration, the values of R^2 and NS were 0.77 and 0.74, respectively, while in the validation, the same performance index values obtained were 0.71 and 0.69, respectively. The results indicate that the model can be accepted for the Upper São Francisco River basin.

CONCLUSIONS

In this study, the SWAT model was used in the Alto São Francisco River basin, Brazil, to simulate streamflow from 1978 to 2007. Modelling studies should always include assessment of uncertainty, and in this study the SUFI-2 technique was used. The calibration and validation results show that the simulated monthly runoff was in reasonable agreement with the measured values, suggesting that the SWAT model could be successfully used to accurately simulate runoff in the study area. The sensitivity analysis of SWAT parameters indicated that runoff is the most sensitive to the curve number and base flow alpha factor. The model overestimates the monthly runoff in the dry periods and underestimates the runoff in the wet periods, but modeling efficiency for monthly runoff was satisfactory. Uncertainty analysis in SUFI-2 indicated that 95PPU could capture more than 80% of the observed streamflow at Pari flow station. The corresponding p-factor and r-factor suggested that there were larger uncertainties at Porto das Andorinhas and Ponte da Taquara. The model also had large uncertainty in peak-flow prediction. These results can be used in further assessments of climate change and land use and cover impact appraisal on water resources for this basin.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to the Brazilian National Research Council (CNPq) and to the Coordination for the Improvement of Higher Education Personnel (CAPES) for their financial support. The authors are also thankful to the Brazilian National Water Agency (ANA) for providing the runoff and rainfall data.

REFERENCES

Abbaspour, K. C. 2007 *User Manual for SWAT-CUP, SWAT Calibration and Uncertainty Analysis Programs*. Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland.

Abbaspour, K., Johnson, C. & van Genuchten, M. T. 2004 Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J.* **3**, 1340–1352.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., de Moraes Gonçalves, J. L. & Sparovek, G. 2013 Köppen's climate classification map for Brazil. *Meteorol. Z.* **22**, 711–728. doi:10.1127/0941-2948/2013/0507.

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**, 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x.

Beskow, S., Mello, C. R., Norton, L. D., Curi, N., Viola, M. R. & Avanzi, J. C. 2009 Soil erosion prediction in the Grande River Basin, Brazil using distributed modelling. *Catena* **79**, 49–59. doi:10.1016/j.catena.2009.05.010.

Beuchle, R., Grecchi, R. C., Shimabukuro, Y. E., Seliger, R., Eva, H. D., Sano, E. & Achard, F. 2015 Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography* **58** (2), 116–127. doi: 10.1016/j.apgeog.2015.01.017.

CEMIG 2015 *2015 São Francisco River Basin*. CEMIG – Energy Company of Minas Gerais State, Belo Horizonte, Brazil. Available in: http://cemig.com.br/en-us/company_and_future/sustainability/programs/environmental_programs/peixe_vivo/river_basins/pages/sao_francisco_river_basin.aspx.

Chen, F. & Li, J. 2016 Quantifying drought and water scarcity: a case study in the Luanhe river basin. *Nat. Hazards* **81**, 1913–1927. doi:10.1007/s11069-016-2162-y.

Coelho, C. A. S., Cardoso, D. H. F. & Firpo, M. A. F. 2016 Precipitation diagnostics of an exceptionally dry event in São Paulo, Brazil. *Theor. Appl. Climatol.* **125**, 769–784. doi:10.1007/s00704-015-1540-9.

Devkota, L. P. & Gyawali, D. R. 2015 Impacts of climate change on hydrological regime and water resources management of the Koshi River Basin, Nepal. *J. Hydrol. Reg. Studies* **4**, 502–515. doi:10.1016/j.ejrh.2015.06.023.

Durães, M. F., Mello, C. R. & Naghettini, M. 2011 Applicability of the SWAT model for hydrologic simulation in Paraopeba River Basin, MG. *Cerne* **17**, 481–488. doi:10.1590/S0104-77602011000400006.

EMBRAPA 2015 *Mapa de solos do Estado de Minas Gerais*. EMBRAPA – Brazilian Agricultural Research Corporation, Brasília, Brazil. http://mapoteca.cnps.embrapa.br.

Fan, F. M., Collischonn, W., Meller, A. B. & Mendes, L. C. 2014 Ensemble streamflow forecasting experiments in a tropical basin: the São Francisco river case study. *J. Hydrol.* **519**, 2906–2919. doi:10.1016/j.jhydrol.2014.04.038.

Fukunaga, D. C., Cecílio, R. A., Zanetti, S. S., Oliveira, L. T. & Caiado, M. A. C. 2015 Application of the SWAT hydrologic model to a tropical watershed at Brazil. *Catena* **125**, 206–213. doi:10.1016/j.catena.2014.10.032.

Grecchi, R. C., Gwyn, Q. H. J., Bénié, G. B., Formaggio, A. R. & Fahl, F. C. 2014 Land use and land cover changes in the

Brazilian Cerrado: a multidisciplinary approach to assess the impacts of agricultural expansion. *Applied Geography* **55**, 300–312. doi:10.1016/j.apgeog.2015.01.017.

IBGE 2016 *Mapa de Biomas*. Available in: http://ftp://geoftp.ibge.gov.br/mapas_tematicos/mapas_murais/biomas.pdf (accessed online 4 December 2016).

Klink, C. A. & Machado, R. B. 2005 Conservation of the Brazilian Cerrado. *Conservation Biology* **19**, 707–713. doi:10.1111/j.1523-1739.2005.00702.x.

Li, L., Zhang, L., Xia, J., Gippel, C. J., Wang, R. & Zeng, S. 2015 Implications of modelled climate and land cover changes on runoff in the middle route of the south to north water transfer project in China. *Water Resour. Manag.* **29**, 2563–2579. doi:10.1007/s11269-015-0957-3.

Lima, J. E. F. W., Montenegro, S. M. G. L., Montenegro, A. A. A. & Koide, S. 2014 Comparative hydrology: relationships among physical characteristics, hydrological behavior, and results of the SWAT model in different regions of Brazil. *Braz. J. Phys. Geogr.* **7**, 1187–1195.

Medrado, E. & Lima, J. E. F. W. 2014 Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. *Geoderma Regional* **1**, 59–66. doi:10.1016/j.geodrs.2014.08.003.

Mello, C. R., Viola, M. R., Norton, L. D., Silva, A. M. & Weimar, A. 2008 Development and application of a simple hydrologic model simulation for a Brazilian headwater basin. *Catena* **75**, 235–247. doi:10.1016/j.catena.2008.07.002.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50**, 885–900. doi: 10.13031/2013.23153.

Narsimlu, B., Gosain, A. K., Chahar, B. R., Singh, S. K. & Srivastava, P. K. 2015 SWAT model calibration and uncertainty analysis for streamflow prediction in the Kunwari River Basin, India, using Sequential Uncertainty Fitting. *Environ. Processes* **2**, 79–95. doi:10.1007/s40710-015-0064-8.

Nash, J. E. & Sutcliffe, J. 1970 River flow forecasting through conceptual models: part I – a discussion of principles. *J. Hydrol.* **10**, 282–290. doi:10.1016/0022-1694(70)90255-6.

Neitsch, S. L., Arnold, J., Kiniry, J., Williams, J. & King, K. 2005 Soil and Water Assessment Tool. Theoretical documentation, version 2005. Texas A&M University, Temple, USA, 618 pp.

Panagopoulos, Y., Makropoulos, C., Baltas, E. & Mimikou, M. 2011 SWAT parameterization for the identification of critical diffuse pollution source areas under data limitations. *Ecol. Modell.* **222**, 3500–3512. doi: 10.1016/j.ecolmodel.2011.08.008.

Patel, D. & Srivastava, P. 2013 Flood hazards mitigation analysis using remote sensing and GIS: correspondence with town planning scheme. *Water Resour. Manag.* **27**, 2353–2368. doi:10.1007/s11269-013-0291-6.

Pereira, D. R., Martinez, M. A., de Almeida André, Q., Pruski, F. F., da Silva, D. D. & Zonta, J. H. 2014 Hydrological simulation using SWAT model in headwater basin in Southeast Brazil. *Eng. Agríc.* **34**, 789–799. doi:10.1590/S0100-69162014000400018.

Pereira, D. R., Martinez, M. A., Pruski, F. F. & da Silva, D. D. 2016a Hydrological simulation in a basin of typical tropical climate and soil using the SWAT model Part I: calibration and validation tests. *J. Hydrol. Reg. Studies* **7**, 14–37. doi:10.1016/j.ejrh.2016.05.002.

Pereira, D. R., Martinez, M. A., da Silva, D. D. & Pruski, F. F. 2016b Hydrological simulation in a basin of typical tropical climate and soil using the SWAT Model Part II: simulation of hydrological variables and soil use scenarios. *J. Hydrol. Reg. Studies* **5**, 149–163. doi:10.1016/j.ejrh.2015.11.008.

Pinto, D. B. F., Silva, A. M., Beskow, S., Mello, C. R. & Coelho, G. 2013 Application of the Soil and Water Assessment Tool (SWAT) for sediment transport simulation at a headwater watershed in Minas Gerais state, Brazil. *Trans. ASABE* **56**, 697–709. doi:10.13031/2013.39200.

Pruski, F. F., Pereira, S. B., Novaes, L. F., Silva, D. D. & Ramos, M. M. 2004 Specific yield discharge and mean precipitation in São Francisco basin. *Rev. Bras. Eng. Agríc. Ambient.* **8**, 247–253. doi:10.1590/S1415-43662004000200013.

Rao, V. B., Franchito, S. H., Santo, C. M. & Gan, M. A. 2016 An update on the rainfall characteristics of Brazil: seasonal variations and trends in 1979-2011. *Int. J. Climatol.* **36**, 291–302. doi:10.1002/joc.4345.

Rodrigues, E. L., Elmiro, M. A. T., Braga, F. A., Jacobi, C. M. & Rossi, F. D. 2015 Impact of changes in land use in the flow of the Pará River Basin, MG. *Rev. Bras. Eng. Agríc. Ambient.* **19**, 70–76. doi:10.1590/1807-1929/agriambi.v19n1p70-76.

Rostamian, R., Jaleh, A., Afyuni, M., Mousavi, S. F., Heidarpour, M., Jalalian, A. & Abbaspour, K. C. 2008 Application of a SWAT model for estimating runoff and sediment in two mountainous basins in central Iran. *Hydrol. Sci. J.* **53**, 977–988. doi:10.1623/hysj.53.5.977.

Sano, E. E., Rosa, R., Brito, J. L. S. & Ferreira, L. G. 2010 Land cover mapping of the tropical savanna region in Brazil. *Environmental Monitoring and Assessment* **166**, 113–124. doi:10.1007/s10661-009-0988-4.

Santana, E. F., Batista, L. V., Silva, R. M. & Santos, C. A. G. 2014 Multispectral image unsupervised segmentation using watershed transformation and cross-entropy minimization in different land use. *GIScience & Remote Sensing* **51**, 613–629. doi:10.1080/15481603.2014.980095.

Santos, C. A. G., Freire, P. K. M. M., Silva, R. M., Arruda, P. M. & Mishra, S. K. 2011 Influence of the catchment discretization on the optimization of runoff-erosion modelling. *Journal of Urban and Environmental Engineering* **5**, 91–102. doi:10.4090/juee.2009.v5n2.091102.

Silva, J. F., Fariñas, M. R., Felfili, J. M. & Klink, C. A. 2006 Spatial heterogeneity, land use and conservation in the Cerrado region of Brazil. *Journal of Biogeography* **33**, 536–548. doi:10.1111/j.1365-2699.2005.01422.xView.

Silva, R. M., Santos, C. A. G., Silva, V. C. L. & Silva, L. P. 2013 Erosivity, surface runoff and soil erosion estimation using GIS-coupled runoff-erosion model in the Mamuaba

Catchment, Brazil. *Environ. Monit. Assess.* **185**, 8977–8990. doi:10.1007/s10661-013-3228-x.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Di Luzio, M. & Srinivasan, R. 2006 A global sensitivity analysis tool for the parameters of multi-variable catchment models. *J. Hydrol.* **324**, 10–23. doi:10.1016/j.jhydrol.2005.09.008.

Versiani, B. R., Carneiro, R. M. F., Amaral, I. R. & Quintão, C. M. F. 2009 Maximum flood regionalization in large basins: study case of the Upper São Francisco region – Minas Gerais, Brazil. *Hydrol. Process.* **23**, 3201–3206. doi:10.1002/hyp.7408.

Wu, H. & Chen, B. 2015 Evaluating uncertainty estimates in distributed hydrological modeling for the Wenjing River watershed in China by GLUE, SUFI-2, and ParaSol methods. *Ecol. Eng.* **76**, 110–121. doi:10.1016/j.ecoleng.2014.05.014.

Xu, Z. X., Pang, J. P., Liu, C. M. & Li, J. Y. 2009 Assessment of runoff and sediment yield in the Miyun Reservoir catchment by using SWAT model. *Hydrol. Process.* **23**, 3619–3630. doi:10.1002/hyp.7475.

Yang, J., Reichert, P., Abbaspour, K., Xia, J. & Yang, H. 2008 Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *J. Hydrol.* **358**, 1–23. doi:10.1016/j.jhydrol.2008.05.012.

First received 1 September 2016; accepted in revised form 2 December 2017. Available online 2 January 2018