

# Rainfall-Runoff Balance Enhanced Model Applied to Tropical Hydrology

## Supplementary Document

Arisvaldo Vieira Mélo Júnior <sup>1</sup>, Lina Maria Osorio Olivos <sup>1</sup>, Camila Billerbeck <sup>1</sup>, Silvana Susko Marcellini <sup>2</sup>, William Dantas Vichete <sup>1,\*</sup>, Daniel Manabe Pasetti <sup>1</sup>, Lígia Monteiro da Silva <sup>2</sup>, Gabriel Anísio dos Santos Soares <sup>2</sup> and João Rafael Bergamaschi Tercini <sup>1</sup>

<sup>1</sup> Departamento de Engenharia Hidráulica e Ambiental, Escola Politécnica, Universidade de São Paulo, São Paulo, 05508-070, Brazil,

arisvaldo@usp.br (A.V.M.J.); lmosorio@usp.br (L.M.O.O.); camila.billerbeck@usp.br (C.B.); daniel.pasetti@usp.br (D.M.P.); tercini@usp.br (J.R.B.T.)

<sup>2</sup> Decision Support System Laboratory (LabSid), Escola Politécnica, Universidade de São Paulo, São Paulo, 05508-070, Brazil, silvana@labsid.eng.br (S.S.M.); ligia@labsid.eng.br (L.M.d.S.); gabriel@labsid.eng.br (G.A.d.S.S.)

**RUBEM**  
Hydrological  
Rainfall rUnoff Balance Enhanced Model QGIS Plugin



**LabSid**

amigos  
da poli 

## Summary

1	INTRODUCTION	2
2	MODEL DEVELOPMENT	2
2.1	Coverage Classification	3
2.2	Model Formulation	4
2.2.1	Water balance	4
2.2.2	Surface runoff	5
2.2.3	Interception	6
2.2.4	Evapotranspiration	8
2.2.5	Lateral flow and groundwater recharge	10
2.2.6	Baseflow	10
2.2.7	Total discharge	10
3	MODEL APPLICATION	11
3.1	Datasets Pre-Processing and Available Sources	12
3.1.1	Soil Data	12
3.1.2	Land Use Data	16
3.1.3	NDVI Data	19
3.1.4	Sources for NDVI Data Collection	19
3.1.5	Gap-Filling in NDVI images - Proposed Methodology	20
3.1.6	Meteorological Data	21
3.2	Dataset Input to Model Application	22
4	MODEL EVALUATION	22
5	MODEL APPLICATION RESULTS	24
6	PLUGIN MODEL INSTALLATION	35

## 1 INTRODUCTION

This supplementary document describes the hydrological process and empirical equations considered by the Rainfall-Runoff Balance Enhanced Model (RUBEM) algorithm. Detailed description of the input dataset, and complementary information regarding input data and its pre-processing requirements are also presented.

RUBEM distributed hydrological model algorithm and its plug-in for QGIS, RUBEM Hydrological, are currently developed under General Public License version 3. The source code repository<sup>1 2</sup> and related documentation<sup>3 4</sup> for both are openly available to the public and are constantly updated. An automatic preprocessing input data application tool is under development. The repository and documentation website are constantly updated.

Chapter 2 present the model development and full Equations used in the RUBEM. Chapter 3 presents the model application. Chapter 4 reference methods applied to model evaluation and the Equations to compute the indicators. Chapter 5 presents the complementary results to the model application, and the last Chapter 6, has the tutorial on model installation into QGIS version 3.x (the LTR 3.16.x version was used in the installation and usage tutorials).

## 2 MODEL DEVELOPMENT

RUBEM is a fully distributed hydrological model that integrates rainfall-runoff processes to represent water balance in soil layers considering interception, evapotranspiration, lateral flows, and aquifer recharge. Water balance process is based on [9, 10, 47]. Figure 1 (main manuscript) shows the structure of hydrological processes used in the model. Figure S1 presents the hydrological physical process considered in the RUBEM algorithm. In the preprocessing task, input data must be spatial into grid cells. Cell grid is defined based on Digital Elevation Models (DEM).

RUBEM integrates classical rainfall-runoff processes, i.e.: interception, evapotranspiration, surface runoff, lateral flows, baseflow, groundwater recharge, and soil water balance. The relation between variables and processes is simulated by the model according to Figure S2. Input data are the Normalized Difference Vegetation Index (NDVI), Land Use Land Cover (LULC), weather, rainfall, and soil, including the DEM. The basin must be divided into grid cells with the spatial resolution adopted by the user, each one with a soil and land cover class. Sub-cell coverage classification is described using percentages of the cell's area covered by fractions of vegetation ( $\alpha_V$ ), bare soil ( $\alpha_S$ ), water ( $\alpha_W$ ) and impermeable ( $\alpha_I$ ).

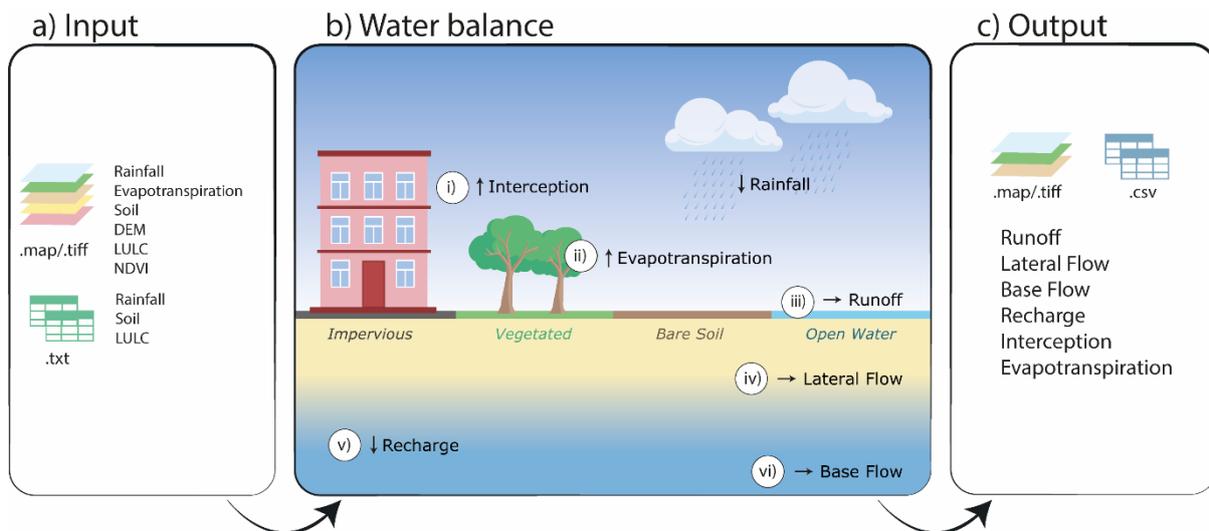
---

<sup>1</sup> Model repository is available at: <https://github.com/LabSid-USP/RUBEM#readme>.

<sup>2</sup> QGIS plug-in repository is available at: <https://github.com/LabSid-USP/RUBEMHydrological#readme>.

<sup>3</sup> Model documentation is available at: <https://rubem.readthedocs.io/en/latest/>.

<sup>4</sup> QGIS plug-in documentation is available at: <https://rubem-hydrological.readthedocs.io/en/latest/>.



**Figure S1.** a) Model inputs, b) water balance calculation, and c) outputs. In b), all the hydrological processes are written in black: rainfall, interception, evapotranspiration, runoff, lateral flow, baseflow, and recharge. In italic and color are the four types of soil used for land cover representation: vegetation ( $\alpha_V$ ), bare soil ( $\alpha_S$ ), water ( $\alpha_W$ ) and impermeable ( $\alpha_I$ )

## 2.1 Coverage Classification

The sub-grid coverage classes and the cell roughness coefficient were defined according to the main cell coverage identification obtained from a previously classified raster image [50]. The sub-grid coverage fractions per coverage type is empirical and do not change during the analysis period, however, it can be modified according to the experience perception of the modeler. Table S1 summarizes the suggested values. The model algorithm associates and assigns at manning roughness parameter according to the corresponding cell coverage type.

**Table S1.** The cover area fraction and roughness coefficients adopted in RUBEM.

Covers Collection 4	Area fraction <sup>1</sup>				Manning roughness <sup>2</sup>
	$\alpha_V$	$\alpha_S$	$\alpha_W$	$\alpha_I$	
Forest Formation	1.000	-	-	-	0.160
Savanna Formation	1.000	-	-	-	0.200
Mangrove	0.700	0.300	-	-	0.150
Forest Plantation	1.000	-	-	-	0.160
Wetland	0.700	0.300	-	-	0.150
Grass Land	0.900	0.100	-	-	0.200
Salt Flat	0.700	0.300	-	-	0.150
Rocky Outcrop	0.200	0.300	-	0.500	0.045
Other Non-Forest Natural Formation	0.900	0.100	-	-	0.200
Pasture	0.800	0.200	-	-	0.150
Agriculture	0.800	0.200	-	-	0.170
Annual and Perennial Crop	0.800	0.200	-	-	0.170
Semi-Perennial Crop	0.800	0.200	-	-	0.170
Mosaic of Agriculture and Pasture	0.800	0.200	-	-	0.160
Beach and Dune	0.300	0.700	-	-	0.040
Urban Infrastructure	0.400	0.100	-	0.500	0.600
Mining	-	1.000	-	-	0.045
Other Non-Vegetated Area	0.400	0.100	-	0.500	0.600
River, Lake, and Ocean	-	-	1.000	-	0.100
Aquaculture	-	-	1.000	-	0.150

1 –[47]; 2 – [46], [86].

## 2.2 Model Formulation

### 2.2.1 Water balance

The soil water balance was estimated in both non-saturated (root zone,  $TU_R$ ) and saturated zones ( $TU_S$ ), shown in Equation S3.  $TU_R$  is related to previous soil moisture content, effectiveness rainfall, and water output (Equation S1). Effective rainfall is given as the difference between rainfall and interception in each cell (Equation S2).  $TU_S$  is related to the previous moisture, the base flow, and the groundwater recharge (Equation S3). A cell fully covered by water fraction sub-grid class is considered saturated, i.e., if  $\alpha_W = 1$  then  $TU_R = TU_S$ .

The root zone water balance Equation adopted in RUBEM aims to obtain the total surface runoff (Equation S1).

$$TU_R = TU_{R,T-1} + P_E - SR - LF - REC - ET_{REAL} \quad S1$$

$$1P_E = P_m - I \quad S2$$

$$TU_S = TU_{S,T-1} - BF + REC \quad S3$$

Where:

$TU_R$  - root zone moisture content (mm);

$TU_{R,t-1}$  - root zone moisture content at the previous time step (mm);

$P_E$  - effective precipitation (mm);

$SR$  - surface runoff (mm);

$LF$  - lateral flow (mm);

$REC$  - groundwater recharge (mm);

$ET_{REAL}$  - total real evapotranspiration (mm);

$P_m$  - total monthly precipitation (mm);

$I$  - total interception (mm);

$TU_S$  - saturated zone moisture content (mm);

$TU_{S,t-1}$  - saturated zone moisture content at the previous time step (mm);

$BF$  - base flow (mm).

### 2.2.2 Surface runoff

The surface runoff is based on the rational method that considers the actual flow ( $C_{SR}$ ) and soil moisture ( $C_h$ ) coefficients (Equation S4).  $C_{SR}$  is obtained using the flow coefficient, related to pervious and impervious areas ( $C_{wp}$ ) adjusted by the average daily rainfall on rainy days ( $\bar{P}_{24}$ ) and the regional consecutive dryness level ( $RCD$ ) (Equation S5).  $C_{wp}$  is calculated by weighting the permeable ( $C_{per}$ ) coefficients and the impermeable surface ( $C_{imp}$ ) (Equation S6). The fractions of water adjust the potential runoff coefficient of the impermeable surface ( $\alpha_w$ ) and impermeable ( $\alpha_I$ ) area (Equations S7 and S8). The permeable area drainage coefficient ( $C_{per}$ ) is obtained from with the weighted sum of land use ( $w1$ ), soil ( $w2$ ), and slope ( $w3$ ).  $w1$ ,  $w2$  and  $w3$  factors must be calibrated (Equation S9).  $C_h$  is related to the soil moisture conditions determined by the root zone soil moisture content and its porosity. The index  $b$  must be calibrated (Equations S10 and S11). The moisture content of the grid is limited to the soil saturation (Equation S12). If the cell is fully covered by water, the surface runoff is given by the difference between rainfall and evapotranspiration (Equation S13).

$$SR = C_{SR} \cdot C_h \cdot (P_m - I) \quad S4$$

$$C_{SR} = \frac{C_{wp} \cdot \bar{P}_{MD}}{C_{wp} \cdot \bar{P}_{MD} - RCD \times C_{wp} + RCD} \quad S5$$

$$C_{wp} = (1 - A_{imp}) \cdot C_{per} + A_{imp} \cdot C_{imp} \quad S6$$

$$C_{imp} = 0.09 \cdot \exp(2.4 \cdot A_{imp}) \quad S7$$

$$A_{imp} = \alpha_0 + \alpha_I \quad S8$$

$$C_{per} = w1 \left( \frac{0.02}{n} \right) + w2 \left( \frac{\theta_{PM}}{1 - \theta_{PM}} \right) + w3 \left( \frac{S}{10 + S} \right) \quad S9$$

$$C_h = \left( \frac{\theta_{TUR}}{\theta_{POR}} \right)^b \quad S10$$

$$\theta_{TUR} = \frac{TU_R}{d_g \cdot Z_r \cdot 10} \quad S11$$

$$\text{If } \theta_{TUR} > \theta_{POR} \text{ then } \theta_{TUR} = \theta_{POR} \quad S12$$

$$\text{If } \alpha_A = 1 \text{ then } SR = P_m - ET_{R,A} \quad S13$$

where:

$SR$  – surface runoff in the cell (mm);

$C_{SR}$  – actual flow coefficient (-);

$Ch$  – coefficient related to soil moisture (-);

$P_m$  – total monthly precipitation (mm);

$I$  – total interception total (mm);

$\bar{P}_{MD}$  – average daily rain on rainy days (mm day<sup>-1</sup> month<sup>-1</sup>);

$RCD$  – regional consecutive dryness factor [parameter to be calibrated] (mm);

$C_{wp}$  – potential runoff coefficient (-);

$C_{imp}$  – potential runoff coefficient of impermeable areas, empirical relationship (-);

$C_{per}$  – potential runoff coefficient of permeable areas (-);

$A_{imp}$  – total fraction of impermeable area per cell (-);

$w_1, w_2, w_3$  – Weights for the three components related to runoff: cover, soil and slope characteristics [parameter to be calibrated] (-);

$n$  – Manning roughness (-);

$\theta_{PM}$  – root layer wilting point volumetric content (-);

$S$  – terrain slope in each cell (%);

$\theta_{TUR}$  – volumetric moisture content of the root layer (-);

$\theta_{POR}$  – volumetric porosity content of the root layer (-);

$b$  – calibration exponent, range between 0 and 1 [parameter to be calibrated] (-);

$TU_R$  – moisture content for the root zone (mm);

$d_g$  – overall soil density in the root layer (g cm<sup>-3</sup>);

$Z_r$  – root layer thickness (cm).

### 2.2.3 Interception

Interception is the fraction of precipitation retained by the canopy of vegetated area. Three vegetation indexes are used to describe and incorporate vegetation characteristics required to obtain the intercepted fraction of precipitation: Fraction of Absorbed Photosynthetically Active Radiation (FPAR), NDVI, and Leaf Area Index (LAI), as shown in Equations S14 to S20.

$$I = \alpha_V \cdot I_V \quad \text{S14}$$

$$I_V = P_m I_R \quad \text{S15}$$

$$I_R = 1 - \exp\left(-\frac{I_D d_p}{P_m}\right) \quad \text{S16}$$

$$I_D = \alpha \cdot LAI \cdot \left(1 - \frac{1}{1 + \frac{P_m [1 - \exp(-0.463 \cdot LAI)]}{\alpha \cdot LAI}}\right) \quad \text{S17}$$

$$LAI = LAI_{max} \cdot \frac{\log(1 - FPAR)}{\log(1 - FPAR_{max})} \quad \text{S18}$$

$$FPAR = \min\left(\frac{(RS - RS_{min}) \cdot (FPAR_{max} - FPAR_{min})}{(RS_{max} - RS_{min})} + FPAR_{min}, 0.95\right) \quad \text{S19}$$

$$RS = \frac{1 + NDVI}{1 - NDVI} \quad \text{S20}$$

where:

$I$  - Interception (mm);

$I_V$  - Interception at the vegetated area (mm);

$\alpha_V$  - Sub-grid vegetated area fraction (%);

$I_R$  - Interception rate, dependent on land cover characteristics, represented by the Leaf Area Index ( $LAI$ ) (mm);

$P_m$  - Total monthly precipitation (mm);

$d_p$  - Number of rainy days in the month (days);

$I_D$  - Minimum threshold for daily interception depends on the canopy storage capacity. Its calculation is associated with the  $LAI$  (mm);

$LAI$  - Leaf Area Index (-);

$\alpha$  - Interception calibration parameter.

$FPAR$  - Fraction Photosynthetically Active Radiation (-);

$FPAR_{min}, FPAR_{max}$  - minimum and maximum values for  $FPAR$  (0.001 and 0.95, respectively), corresponding to the minimum and maximum values for  $LAI$  for a particular vegetation class.

$RS$  - Factor related to NDVI.

LAI is estimated using FPAR derived from NDVI satellite images, according to Equations S18 to S20 [47, 86].

#### 2.2.4 Evapotranspiration

Evapotranspiration refers to the transfer of water from the soil-plant system to the atmosphere. Actual evapotranspiration is obtained using the sum of evapotranspiration values for each sub-grid cover fraction (vegetation, bare soil, water, and impervious soil) (Equations S21, S22, S25, S28, and S30). The sub-grid evapotranspiration fraction is related to the potential evapotranspiration ( $ET_p$ ) – [47], required as input data. In this study,  $ET_p$  is previously calculated using the Penman-Montheit method [49]. The crop coefficient ( $kc$ ) and the coefficient of moisture reduction ( $ks$ ) are calculated from  $NDVI$  and soil moisture content (Equations S23, S24, S26, and S27).  $ET_{R,W}$  depends on the water evaporation coefficient (Equation S29), and  $ET_{R,I}$  is equal to intercept loss (Equation S30).

$$ET_{REAL} = \alpha_V ET_{R,V} + \alpha_S ET_{R,S} + \alpha_W ET_{R,W} + \alpha_I ET_{R,I} \quad S21$$

$$ET_{R,V} = ET_P \cdot kc \cdot ks \quad S22$$

$$kc = kc_{min} + (kc_{max} - kc_{min}) \cdot \left( \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right) \quad S23$$

$$\text{If } NDVI \leq 1.1 \cdot NDVI_{min} \text{ then } kc = kc_{min} \quad S24$$

$$ET_{R,S} = ET_P \cdot kc_{min} \cdot ks \quad S25$$

$$ks = \frac{\ln(TU_R - TU_{PM} + 1)}{\ln(TU_{CC} - TU_{PM} + 1)} \quad S26$$

$$\text{If } TU_R < TU_{PM} \text{ then } ks = 0 \quad S27$$

$$ET_{R,W} = \frac{ET_P}{kp} \quad S28$$

$$kp = 0.482 + 0.024 \cdot \ln(B) - 0.000376U_2 + 0.0045UR \quad S29$$

$$ET_{R,I} = I \quad S30$$

Where:

$ET_{REAL}$  – real evapotranspiration (mm);

$\alpha_V$  – vegetated fraction of cell area (%);

$\alpha_S$  – bare soil fraction of cell area (%);  
 $\alpha_W$  – water fraction of cell area (%);  
 $\alpha_I$  – impermeable fraction of cell area (%);  
 $ET_{R,V}$  – real evapotranspiration of the vegetated area (mm);  
 $ET_{R,S}$  – real evapotranspiration of the bare soil area (mm);  
 $ET_{R,W}$  – real evapotranspiration of the water area (mm);  
 $ET_{R,I}$  – real evapotranspiration of the impermeable area (mm);  
 $ET_P$  – potential evapotranspiration (mm);  
 $kc$  – crop coefficient (-);  
 $kc_{max}$  e  $kc_{min}$  – maximum and minimum possible values for crop coefficient;  
 $ks$  – soil moisture reduction coefficient (-);  
 $NDVI_{max}$  e  $NDVI_{min}$  – maximum and minimum values of the standardized vegetation index obtained based on the historical NDVI series for each cell;  
 $TU_R$  – root zone moisture content (mm);  
 $TU_{PM}$  – wilting point moisture content in the root zone (mm);  
 $TU_{CC}$  – field capacity moisture content in the root zone (mm);  
 $kp$  – water evaporation coefficient (-);  
 $B$  – Class A Tank Border Width [between 20 to 30 m] (m);  
 $U_2$  – average wind speed at 2 m above ground surface (m/s);  
 $UR$  – relative humidity (%);  
 $I$  – Interception [from 1 to 3 mm];

The vegetated area evapotranspiration ( $ET_{R,V}$ ) was calculated using Equations S22 S23, and S26. The crop coefficient ( $kc$ ) depends on the maximum and minimum possible values range for the crop coefficient ( $kc_{max}$  e  $kc_{min}$ ) and of NDVI index, subject to: if  $NDVI \leq 1.1 \cdot NDVI_{min}$ ,  $kc = kc_{min}$ . The soil moisture reduction coefficient ( $ks$ ) relies on the moisture content. If the moisture content of the root area ( $TU_R$  - mm) is lower than the moisture content of the soil at the wilting point ( $TU_{PM}$  - mm), the coefficient of moisture reduction ( $ks$ ) is zero.  $ET_{R,S}$  is obtained using  $kc_{min}$ .  $ET_{R,A}$  is obtained using the evaporation coefficient  $kp$  (Equation S29), and  $ET_{R,I}$  is based on losses caused by interception.

### 2.2.5 Lateral flow and groundwater recharge

Lateral flow occurs in the soil unsaturated layer. Groundwater recharge is the downflow from the surface to the water aquifer through infiltration and deep percolation. The lateral flow ( $LF$ ) and recharge of saturated zone ( $REC$ ) are based on the root layer moisture content, hydraulic conductivity ( $K_R$ ), and vertical and horizontal flow partitioning coefficients ( $f$ ), Equations S31 and S32.

$$LF = f \cdot K_R \left( \frac{TU_R}{TU_{SAT}} \right)^2 \quad S31$$

$$REC = (1 - f) \cdot K_R \cdot \left( \frac{TU_R}{TU_{SAT}} \right)^2 \quad S32$$

Where:

$LF$  – lateral flow (mm);

$REC$  – recharge (mm);

$f$  – vertical and horizontal flow partitioning coefficient [parameter to be calibrated] (-);

$K_R$  – root zone hydraulic conductivity coefficient (mm month<sup>-1</sup>);

$TU_R$  – root zone moisture content (mm);

$TU_{SAT}$  – root zone moisture content at saturation (mm).

### 2.2.6 Baseflow

Baseflow is calculated according to [10]. Baseflow is related to groundwater recharge and recession coefficient, and its calculation is subject to a specific threshold attributed for each basin (Equation S33).

$$BF = \begin{cases} 0, & \text{if } TU_S \leq BF_{thresh} \\ BF_{t-1} \cdot e^{-\alpha_{gw}} + (1 - e^{-\alpha_{gw}}) \cdot REC, & \text{if } TU_S > BF_{thresh} \end{cases} \quad S33$$

Where:

$BF$  - baseflow (mm);

$BF_{t-1}$  - baseflow at the previous time step (mm);

$\alpha_{gw}$  - base flow decay coefficient [parameter to be calibrated] (-);

$REC$  - recharge (mm);

$TU_S$  - saturated zone moisture content (mm);

$BF_{thresh}$  - threshold baseflow, attributed for each watershed (mm).

### 2.2.7 Total discharge

Total discharge ( $Q_{Tot}$ ) corresponds to the total amount of surface runoff ( $SR$ ), lateral flow ( $LF$ ) and base flow ( $BF$ ) produced in each cell, as showed by Equation S34. The total superficial flow is computed at any grid cell, and the total flow is computed using an aggregate

result according to the LDD [10,53]. The total discharge at the cell level is calculated by summarizing the discharge of the previous month and the discharge of the current month adjusted by the damping coefficient ( $x$ ), Equation S35. The total discharge in cell ( $Q_t$ ) is determined by the product of the area in square meters divided by the number of seconds in a month.

$$Q_{Tot} = SR + LF + BF \quad S34$$

$$Q_t = x \cdot Q_{t-1} + \frac{0.001 \cdot (1 - x) \cdot A \cdot Q_{Tot}}{days \cdot 24 \cdot 3600} \quad S35$$

Where:

$Q_{Tot}$  – total discharge in cell (mm);

$SR$  – surface runoff (mm);

$LF$  – lateral flow (mm);

$BF$  – base flow (mm);

$Q_t$  – total discharge in cell ( $m^3 s^{-1}$ );

$Q_{t-1}$  – total discharge in cell at the previous time step ( $m^3 s^{-1}$ );

$A$  – cell area ( $m^2$ );

$x$  – damping coefficient [parameter to be calibrated] (-).

### 3 MODEL APPLICATION

Three Brazilian basins were selected for RUBEM model application and testing, two of which are in wet regions (Piracicaba River Basin - PRB and Upper Iguaçu River Basin – UIRB). The third basin has two well defined hydrological zones, a dry climate in the western region, and a wet climate in the eastern portion (Ipojuca River Basin - IRB). Figure 3 of manuscript shows the selected areas. The selected basins represent distinct characteristics of LULC, soil, evaporation, and hydrological regions. The regions also distinct in terms of water use demands, with primary needs for public supplies, industry, and agriculture, as shown in Table S7.

The IRB is located between the parallel 8° '09' and 8° '40' south latitude, and the meridians 34° '57' and 37° '02' west longitude. The basin area of 3,433.5 km<sup>2</sup> covers 24 municipalities. Due to its elongated conformation in the west-east direction, the upper and middle portion of the basin are in the arid and semiarid regions. The lower portion is predominantly located in the rainforest zone, including the coastal strip. IRB climate presents tropical rainy monsoons with dry summer. The average annual rainfall is 640 mm upstream, in the intermediate course of 795 mm and 2267 mm near the coast. The annual potential evapotranspiration is 1875 mm upstream part and in the middle course of the basin and decreases to the coast to 1550 mm. The average annual flows range from 0.0 to 2.0 L s<sup>-1</sup> km<sup>-2</sup> in the dry region and increase to the coast. Land use cover classification data shows IRB is covered by pastures (49%), wooded and deciduous forests (28.7%), surrounded by agricultural land (13.5%), with old secondary growth forest fragments (4.7%), urban infrastructure (3.2%), rivers and artificial water reservoir (0.11%).

The PRB is located between parallel 22° and 23° '30' south latitude, and the meridians 46° and 49° west longitude, covering an area of 10,701.38 km<sup>2</sup>, with 89 municipalities. The basin is located in the subtropical wet climate region with dry winters (April to September) and wet summers (October to March). The average annual rainfall averages are between 1,195 mm and 1,609 mm. The annual potential evapotranspiration is lower in eastern region (438 mm) and higher in western portion (976 mm). The average flow at the basin is 13.2 L s<sup>-1</sup> km<sup>-2</sup> [88]. Cultivated forests cover 20.8% of the basin area, followed by agriculture and pasture mosaics (15.6%), natural prairies (12.8%), urban infrastructure (10.7%), natural forests (9.5%), rivers, and artificial water reservoirs (1.2%).

The UIRB is located between parallel 25° '13' and 25° '50' south latitude, and meridians 48° 57' 11" and 49° 36' 22" west longitude, covering an area of 2,695.8 km<sup>2</sup>, with 18 municipalities. The climate of the basin is humid subtropical mesothermic, without dry season, mild summers, and cold winters. Annual rainfall averages range from 1,338 mm to 1,801 mm, declining from east to west and north to south due to the influence of mountain formations. The average annual evapotranspiration for 2006, 2007, and 2009 was 1054 mm. The average flow is 19.25 L s<sup>-1</sup> km<sup>-2</sup> [89]. UIRB main land use and cover types are pastures and agriculture (42.1%), followed by natural forests (33%), urban infrastructure (21.5%), cultivated forests (1.5%), rivers, and lakes (1.5%).

### 3.1 Datasets Pre-Processing and Available Sources

RUBEM model is written in Python programming language, using the PCRaster framework [53] into a GIS<sup>5</sup> environment. The input dataset must be pre-processed using the PCRaster time-series data class on the model domain.

The RUBEM documentation provides specific details and requirements of the input dataset. Useful time-series related algorithms: land user and land cover, precipitation, NVDI, soil map, evapotranspiration, and other pre-processing scripts are available (see the documentation for more information<sup>6</sup>).

This chapter presents the input data sources used for the selected basin model application, and its pre-processing details.

#### 3.1.1 Soil Data

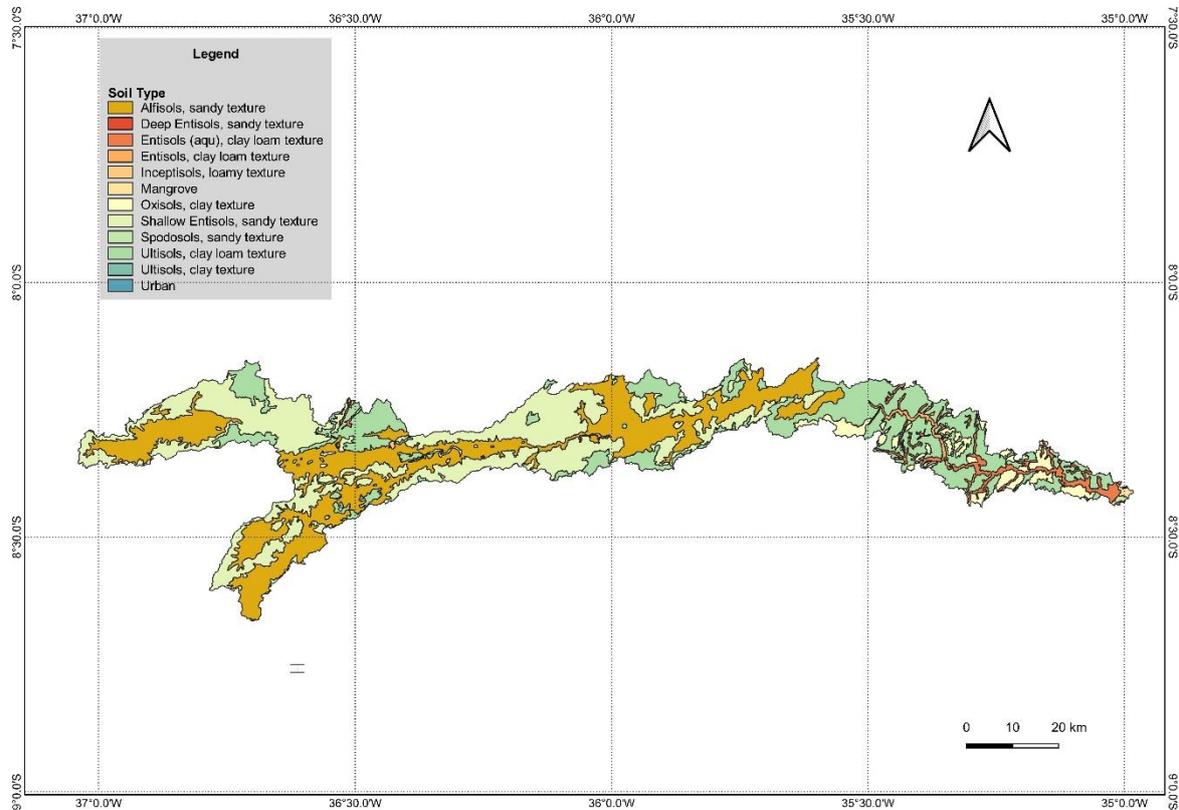
RUBEM requires soil data in the form of lookup tables related to a map of soil class. The different soil types for each basin were established from a reclassification of available semi-detailed maps according to the Brazil Soil Map [90]. The differentiation of the soils considering their depth and textural class (from semi-detailed surveys) greatly influences their hydrological behavior. The reclassification also sought to simplify the number of soil types, as the great diversity of associations in the surveys made it impossible to calibrate the model. The reclassifications were specific to each basin according to data availability and soil diversity. Figures S2, S3, and S4 present the final map results.

---

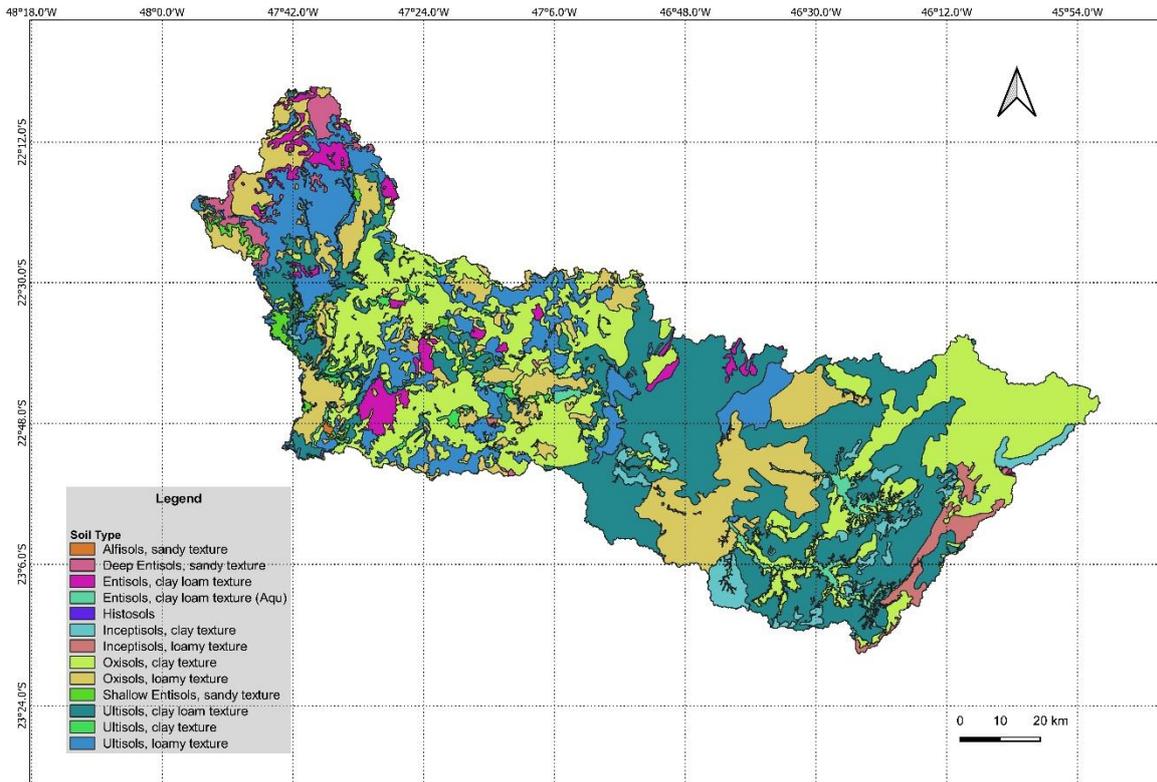
<sup>5</sup> Geographic Information System (QGIS was used as the GIS environment and the model's functionality was added as a plug-in for it.).

<sup>6</sup> <https://rubem-hydrological.readthedocs.io/en/latest/preprocessing.html>

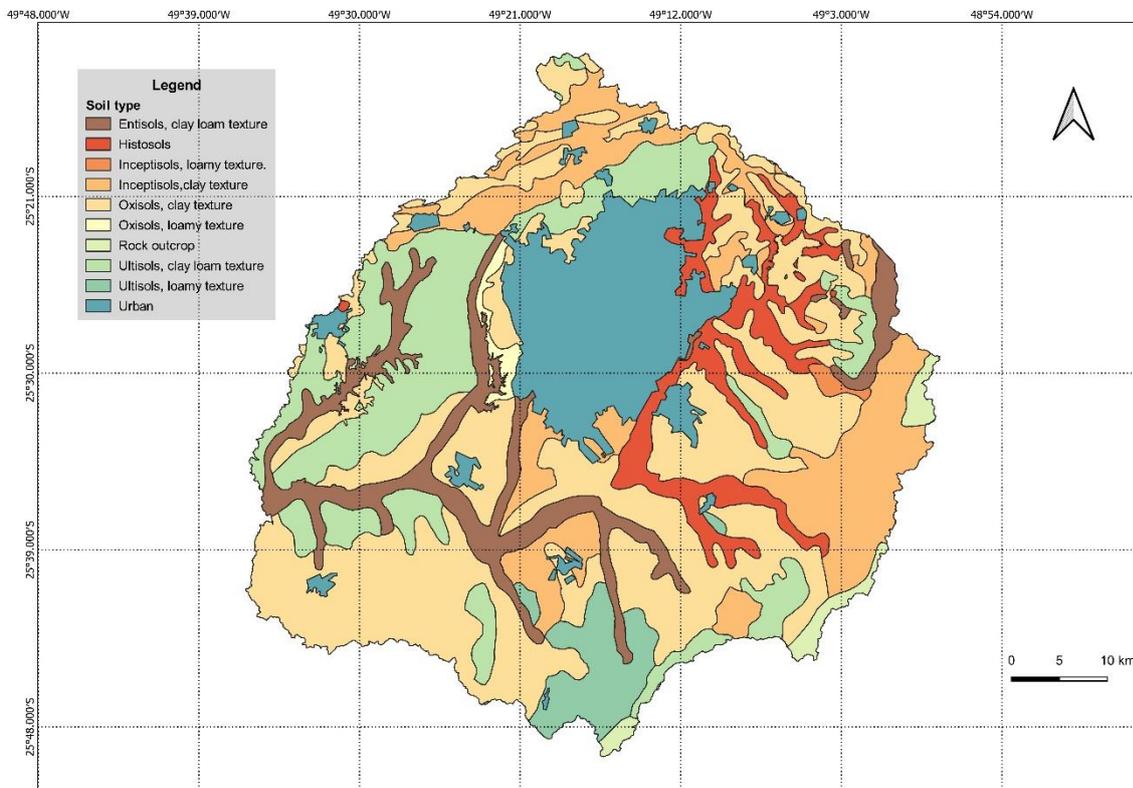
Table S2 shows the summary of the soils found in the three study areas and the associated parameters. The primary data source was the HYBRAS database [54], a hydro-physical database of soils in Brazil, with data on water retention and saturated hydraulic conductivity. Other references to obtain averages for each parameter were also included.



**Figure S2.** Ipojuca River Basin RUBEM Soil Map classification.



**Figure S3.** Piracicaba River Basin RUBEM Soil Map classification.



**Figure S4.** Upper Iguazu River Basin RUBEM Soil Map classification.

**Table S2.** Parameters adopted for soil classes.

Soil classification <sup>1</sup>	K <sub>R</sub> <sup>2</sup>	D <sub>g</sub> <sup>3</sup>	Z <sub>r</sub> <sup>4</sup>	θ <sub>SAT</sub> <sup>5</sup>	θ <sub>fc</sub> <sup>6</sup>	θ <sub>w</sub> <sup>7</sup>	References
Ultisols, clay loam texture	153.16	1.54	150.39	0.46	0.26	0.19	[91-96]
Inceptisols, loamy	235.71	1.31	137.45	0.55	0.29	0.18	[93,95,97-99]
Spodosols, sandy	253.91	1.41	131.25	0.47	0.22	0.08	[54]
Entisols, clay loam	89.35	1.54	114.30	0.42	0.31	0.20	[98,100]
Oxisols, clay texture	237.27	1.25	200.81	0.58	0.36	0.27	[94,99,101-104]
Deep Entisols, sandy texture	367.88	1.59	200.00	0.33	0.08	0.03	[105-110]
Shallow Entisols, Sandy texture	367.88	1.59	60.00	0.33	0.08	0.03	[93,111,112]
Entisols, clay loam texture	153.16	1.46	120.00	0.43	0.28	0.22	[99,113]
Ultisols, clay texture	164.16	1.38	177.20	0.52	0.38	0.28	[99,111,112,114]
Alfisols, sandy texture	367.88	1.64	148.07	0.38	0.29	0.17	[93,100,115-117]
Mangrove	367.88	1.31	120.00	0.33	0.12	0.06	[118]
Ultisols, loamy texture	352.39	1.51	147.50	0.47	0.22	0.14	[95,96,99,119-121]
Inceptisols, clay texture	267.69	1.18	132.52	0.54	0.37	0.22	[95,97-99]
Oxisols, loamy texture	187.77	1.31	189.77	0.45	0.26	0.15	[99,101,102,122-124]
Histosols	89.35	1.54	50.00	0.42	0.31	0.20	[125]
Urban Area IRB	203.53	1.48	140.82	0.47	0.24	0.13	Nearest neighbor Mean
Urban Area UIB	252.48	1.22	166.67	0.56	0.37	0.24	Nearest neighbor Mean

<sup>1</sup>Relation between Brazilian Soil Classification and US soil taxonomy: Argissolos (Ultisols), Cambissolos (Inceptisols), Latossolos (Oxisols), Organossolos (Histosols), Neossolos (Entisols quartzipsamment), Espodossolos (Spodosols), Gleissolos (Entisols -Aqu-alf-and-ent-ept); <sup>2</sup>Hydraulic Conductivity [mm/month]; <sup>3</sup>Soil Bulk Density [g/cm<sup>3</sup>]; <sup>4</sup>Root-zone Depth [cm]; <sup>5</sup>Saturated capacity water content [θ (cm<sup>3</sup>/cm<sup>3</sup>)]; <sup>6</sup>Wilting Point water content [θ (cm<sup>3</sup>/cm<sup>3</sup>)]; <sup>7</sup>Field capacity water content [θ (cm<sup>3</sup>/cm<sup>3</sup>)].

### 3.1.2 Land Use Data

Land use and cover data is fundamental for basin characterization. RUBEM requires this information as input to obtain interception, evapotranspiration, runoff and assign parameters such as Manning roughness.

Due to the size of the domain area of the study and the complexity of the coverage classification procedures, it was taken into consideration the database available in the Annual Mapping of Land Cover and Land Use Project in Brazil (MapBiomias) - Collection 4 [50].

MapBiomias images are based on machine learning algorithms for classification of land cover, resulting in to 27 classes and subclasses, for Collection 4 products, as presented in Table S3.

**Table S3.** Land Use and Land Cover Classification – MapBiomias [50].

<b>Collection 4</b>	<b>New ID</b>
<b>1. Forest</b>	<b>1</b>
1.1. Natural Forest	2
1.1.1. Forest Formation	3
1.1.2. Savanna Formation	4
1.1.3. Mangrove	5
1.2. Forest Plantation	9
<b>2. Non-Forest Natural Formation</b>	<b>10</b>
2.1. Wetland	11
2.2. Grassland	12
2.3. Salt Flat	32
2.4. Rocky Outcrop	29
2.5. Other Non-Forest Natural Formation	13
<b>3. Farming</b>	<b>14</b>
3.1. Pasture	15
3.2. Agriculture	18
3.2.1. Annual and Perennial Crop	19
3.2.2. Semi-perennial Crop	20
3.3. Mosaic of Agriculture and Pasture	21
<b>4. No vegetated area</b>	<b>22</b>
4.1. Beach and Dune	23
4.2. Urban Infrastructure	24
4.3. Mining	30
4.4. Other Non-Vegetated Area	25
<b>5. Water</b>	<b>26</b>
5.1. River, Lake, and Ocean	33
5.2. Aquaculture	31
<b>6. Non Observed</b>	<b>27</b>

The characterization and analysis of the historical series of watershed coverage (LULC) and the assignment of modeling parameters involved the following steps:

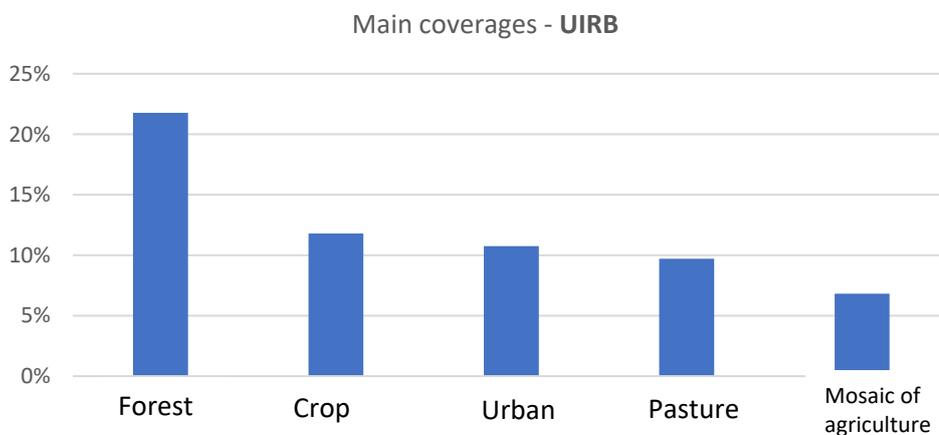
1. Download the series of coverages for the period 2000 – 2018 from Mapbiomas (collection 4) through the GEE platform;
2. Preparation of files to PCRaster format and envelopes of interest;
3. Assignment of features for modeling (area fractions per pixel and Manning roughness).

Steps 2 and 3 were performed using algorithms implemented in Python to automate the analysis. A brief characterization of the LULC of the study area is presented in the following pages (Figures S5a, S5b, and S5c).

### Upper Iguazu River Basin



- 11 classes of coverage
- Forest formation predominates (average of 21.8% of the area in the 1985-2018 series)



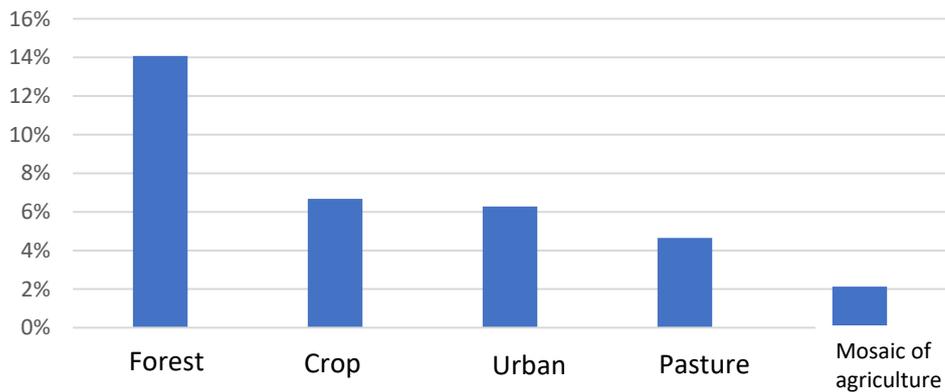
**Figure S5a.** Summary of coverage analysis for the Upper Iguazu River Basin.

### Piracicaba River Basin



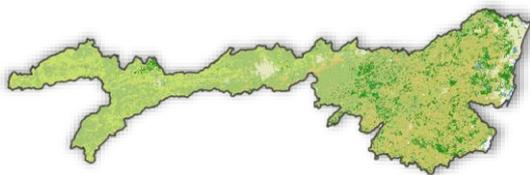
- 14 classes of coverage
- Pasture predominates (average of 14.1% of the area in the series of 1985-2018)

Main coverages - PRB



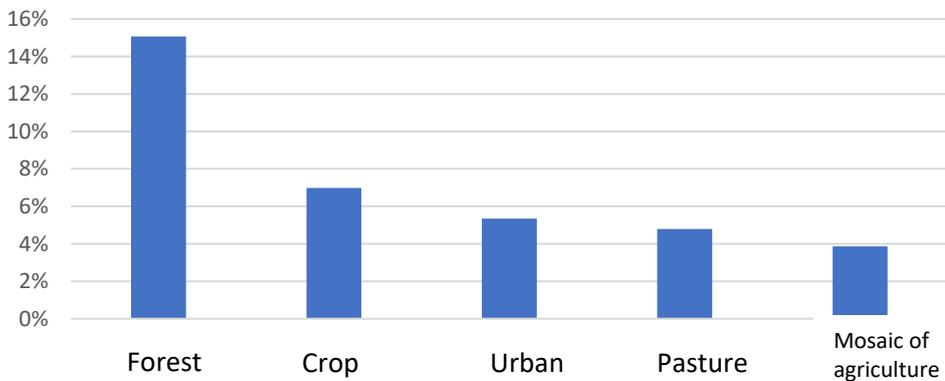
**Figure S5b.** Summary of coverage analysis for the Piracicaba River Basin.

### Ipojuca River Basin



- A more significant number of coverages (16 classes)
- Natural forest closer to the coast and Savanna Formation in center region.
- Average pasture around 5% of the area between 1985-2018

Main coverages - IRB



**Figure S5c.** Summary of coverage analysis for the Ipojuca River Basin.

### 3.1.3 NDVI Data

The main benefit of using remote sensing data is the access to physical and biophysical parameters that describe natural elements, especially vegetation. Vegetation indicators are dimensionless radiometric measurements that indicate the relative abundance and activity of vegetation. Among the most used vegetation indicators, the Leaf area index (LAI), absorbed photosynthetically active radiation (Fraction of Absorbed Photosynthetically Active Radiation - FPAR), and the Normalized Difference Vegetation Index – NDVI are used to model the water balance [126].

NDVI describes the normalized ratio between the near-infrared and red bands, formally described according to Equation S43.

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad S43$$

Where  $\rho_{nir}$  corresponds to the value of the near-infrared band and  $\rho_{red}$  to the value of the red band of the electromagnetic radiation spectrum. The NDVI index allows monitoring seasonal and interannual changes in vegetation development and activity, being widely used in environmental and natural science research [126].

NDVI ranges from -1 to 1. The closer to 1, the greater the vegetative activity in the location represented by the pixel, while negative values or close to 0 indicate areas of water, buildings, bare soil, in short, where it is little or no chlorophyll activity.

### 3.1.4 Sources for NDVI Data Collection

NDVI data can be obtained using specific remote sensing products. The Landsat and MODIS (Moderate Resolution Imaging Spectroradiometer) sensors are the most commonly used sources. Landsat is a joint program of the USGS and NASA, and their satellites have been collecting data since 1972. Landsat satellites represent the Earth's entire surface at a resolution of 30 m, including multispectral and thermal data. The calculations to obtain the NDVI can be performed with data from treated images. MODIS started in 2000 and is the longest continuous daily record of Earth satellite observation ever compiled. The MODIS sensor is present on the Terra and Aqua satellites, and its best resolution is 250 m. The MODIS sensor database has products such as NDVI already processed and available.

A quantitative study was carried out with the aid of the Google Earth Engine (GEE) platform to analyze the availability of images from the Landsat and Terra satellites. In this selection of images, treatment was performed to mask pixels with clouds, which do not have data information. LANDSAT images have a band that informs the percentage of clouds in the image (Cloud Cover), being selected images with up to 50% of clouds. The MODIS sensor images have pixel information indicating the same quality, with low-quality pixels possibly representing clouds, and were removed.

Based on the analysis of image availability, the product MODIS - MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250 m, was selected to obtain NDVI data. NDVI is generated every 16 days through a mosaic of daily observations. This process ensures that at

least two NDVI observations for the areas of interest are available monthly. Images collected presented gaps in some areas, where gap-filling treatment was provided.

### 3.1.5 Gap-Filling in NDVI images - Proposed Methodology

A simplified methodology for NDVI gap-filling was developed to obtain full monthly images with consistent information linked to annual coverage maps. This association proved coherent. Even if the frequency of coverage is annual, the reading of NDVI monthly reflects the characteristics and state of maturity of the vegetation cover.

The developed methodology consists of 4 steps:

#### **Step 1: NDVI image pairing with faults and associated coverage map**

In this step, the monthly NDVI image is associated with the corresponding annual coverage map.

#### **Step 2: obtaining the average NDVI values by coverage class**

From the valid or flawless pixels of the NDVI image, a reading associated with each coverage class of the annual map was taken to obtain the average NDVI values for the month of analysis.

#### **Step 3 – Definition of filling criteria**

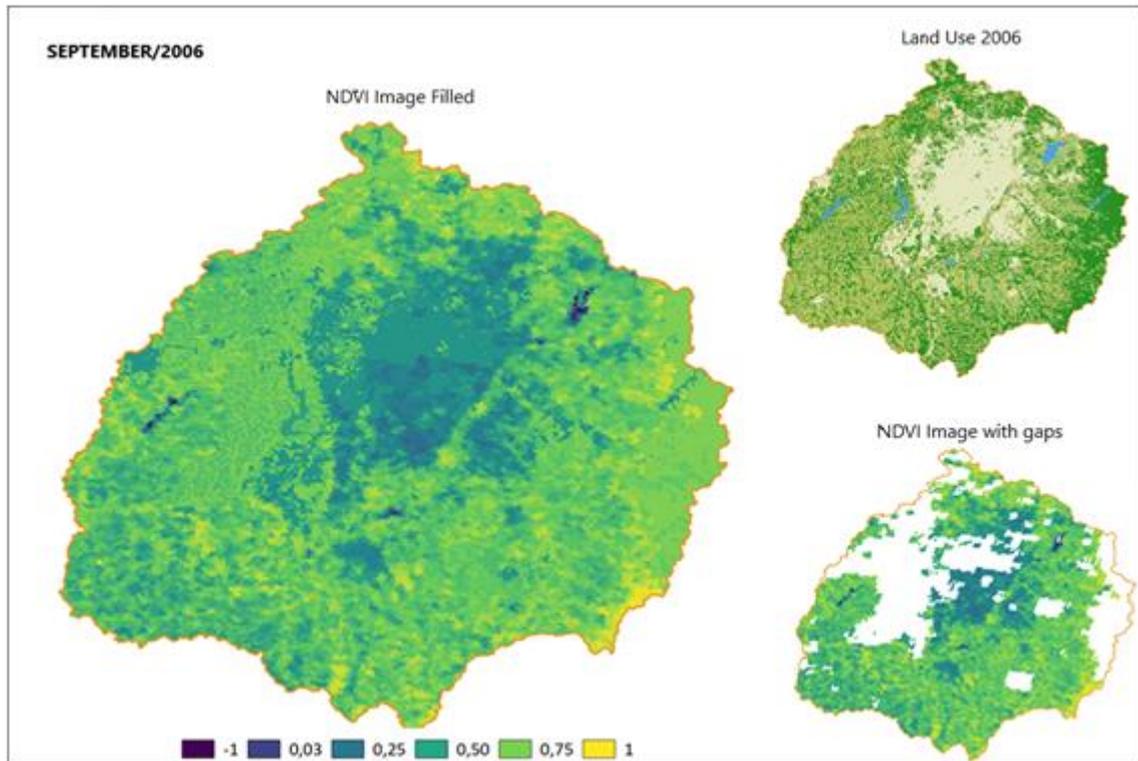
The criterion for gap-filling in faulty pixels is selected after compiling the monthly average NDVI values for each coverage class. Three criteria were listed for filling gaps, which depend on flawless pixel and coverage availability, in the month under analysis. The criteria are follows:

- **Criteria 1** – filling pixels with flaws associated with the average NDVI values obtained in the same month, by coverage class;
- **Criteria 2** – completion with mean NDVI values obtained in the immediately preceding and subsequent months, by coverage class;
- **Criteria 3** – completion with annual average NDVI values, by coverage class;

#### **Step 4 – Filling the Gaps**

Finally, monthly NDVI gaps were filled in according to the criterion selected for filling in a group of pixels associated with a given coverage class.

Figure S6 shows an example of the result obtained with the described methodology, for the image from September/2006 of Upper Iguacu River Basin.



**Figure S6.** Example of result for filling the NVDI of the image referring to the month 9/2006 of Upper Iguacu River Basin.

### 3.1.6 Meteorological Data

The meteorological data necessary to run the model is summarized in:

- Rainfall
- Potential Evapotranspiration
- Class A Pan Coefficient ( $K_p$ )

Rainfall was obtained from stations distributed in each basin area. For potential evapotranspiration, point values (at station locations) were calculated using the Penman-Montheit method [49], and the Class A pan coefficient was calculated using the wind speed, and relative humidity. In all cases, the point values were spatialized using the kriging interpolation method. Parameters of the variograms and the conversion to the PCRaster format were generated using Python scripts that can be found at <https://rubem-hydrological.readthedocs.io/en/latest/preprocessing.html>.

### 3.2 Dataset Input to Model Application

The dataset used in the RUBEM application is available from Hydroshare.

Méllo, A. V., L. M. O. Olivos, C. Billerbeck, S. S. Marcellini, W. D. Vichete, D. M. Pasetti, L. M. d. Silva, G. A. d. S. Soares, J. R. B. Tercini (2022). *Rainfall-Runoff Balance Enhanced Model Applied to Tropical Hydrology*, HydroShare, <https://doi.org/10.4211/hs.6f3670b8cd944e7ea72e03d1b9ca928f>, DOI for this published resource is pending activation.

A video on the LabSid YouTube channel provides instructions for using the Ipojuca River Basin (IRB) input dataset (<https://youtu.be/R8CcLSkLj0Q>). In addition, the RUBEM Hydrological documentation provides a tutorial for model application using the Upper Iguaçú River Basin (UIRB) input dataset (<https://rubem-hydrological.readthedocs.io/en/latest/tutorials.html>).

All raster maps of the datasets that were used to apply the RUBEM model (also including the respective output data generated) have the following characteristics:

- Spatial Resolution: 500m
- Temporal Resolution: monthly
- Period time: January/2000 - December/2018
- Available data: meteorological forcings (precipitation and evapotranspiration), hydrologic budget terms (evapotranspiration, runoff, soil moisture storage, and total storage) and maps of gridded parameters (land mask, soil layers, total storage capacity, and NDVI).

## 4 MODEL EVALUATION

The current version of the RUBEM does not feature an automatic calibration process, however, user can calibrate the parameters manually by comparing model output with expected results of hydrological variables. The differential evolution algorithm adopted for the automatic calibration process for model application was introduced for this purpose by [55]. This algorithm can be integrated into the RUBEM in future versions. The selection criteria include a series with a few flags in the basin according to the observed data at gauges stations, LULC classification, and rainfall regimes. The objective function of the automatic calibration algorithm was the Nash-Sutcliffe efficiency index.

These parameters are obtained during the calibration period and then tested during the validation period. In both periods, model performance is measured using indicators (Table S4). Root median square error – *RMSE* [62], Nash-Sutcliffe Efficiency – *NSE* [56], relative bias – *RB* [59], and asynchronous regression method [60] are used as performance indicators to evaluate the model application results (S36 to S42).

The best group of parameters is revealed through three objective functions: Root n Square Error – *RMSE*, Equation S36 [62], number of times that the variability of the observations is greater than the mean error –  $n_t$  and Nash-Sutcliffe Efficiency – *NSE*, Equations S37 and S38 [56], relative bias - *RB*, Equation S39 [59], and the asynchronous

regression method that determines the function  $F(Q_s) = u \cdot F(Q_o)$  according to the percentiles of  $q_o$  and  $q_s$  of the distribution of  $F(Q_o)$  and  $F(Q_s)$  at each probability level, Equations from S40 to S42 [60].

$$RMSE = \sqrt{\frac{\sum_1^N (Q_s - Q_o)^2}{N}} \quad SE 36$$

$$n_t = \frac{SD}{RMSE} - 1 \quad SE 37$$

$$NSE = 1 - \left( \frac{1}{n_t + 1} \right)^2 \quad SE 38$$

$$RB = \frac{\frac{1}{N} \cdot \sum_1^N (Q_s - Q_o)}{Q_o} \quad SE 39$$

$$F(Q_s) = u \cdot F(Q_o) \quad SE 40$$

$$F(Q_s) = \frac{i}{N_s} \quad SE 41$$

$$F(Q_o) = \frac{i}{N_o} \quad SE 42$$

Where:

$RMSE$  – Root Mean Square Error (-);

$SD$  – standard deviation of observed data;

$n_t$  – number of times that the observations variability is greater than the mean error;

$NSE$  – Nash-Sutcliffe Efficiency (-);

$RB$  – relative bias (-);

$Q_o$  – observed discharge ( $m^3 s^{-1}$ );

$Q_s$  – simulated discharge ( $m^3 s^{-1}$ );

$N$  – sample size;

$N_o$  – sample size of observed discharge;

$N_s$  – sample size of simulated discharge;

$F(Q_o)$  – cumulative distribution function of the percentile of  $Q_o$ ;

$F(Q_s)$  – cumulative distribution function of the percentile of  $Q_s$

**Table S4.** Criteria for assessing the performance of the model based on efficiency indexes [56,59].

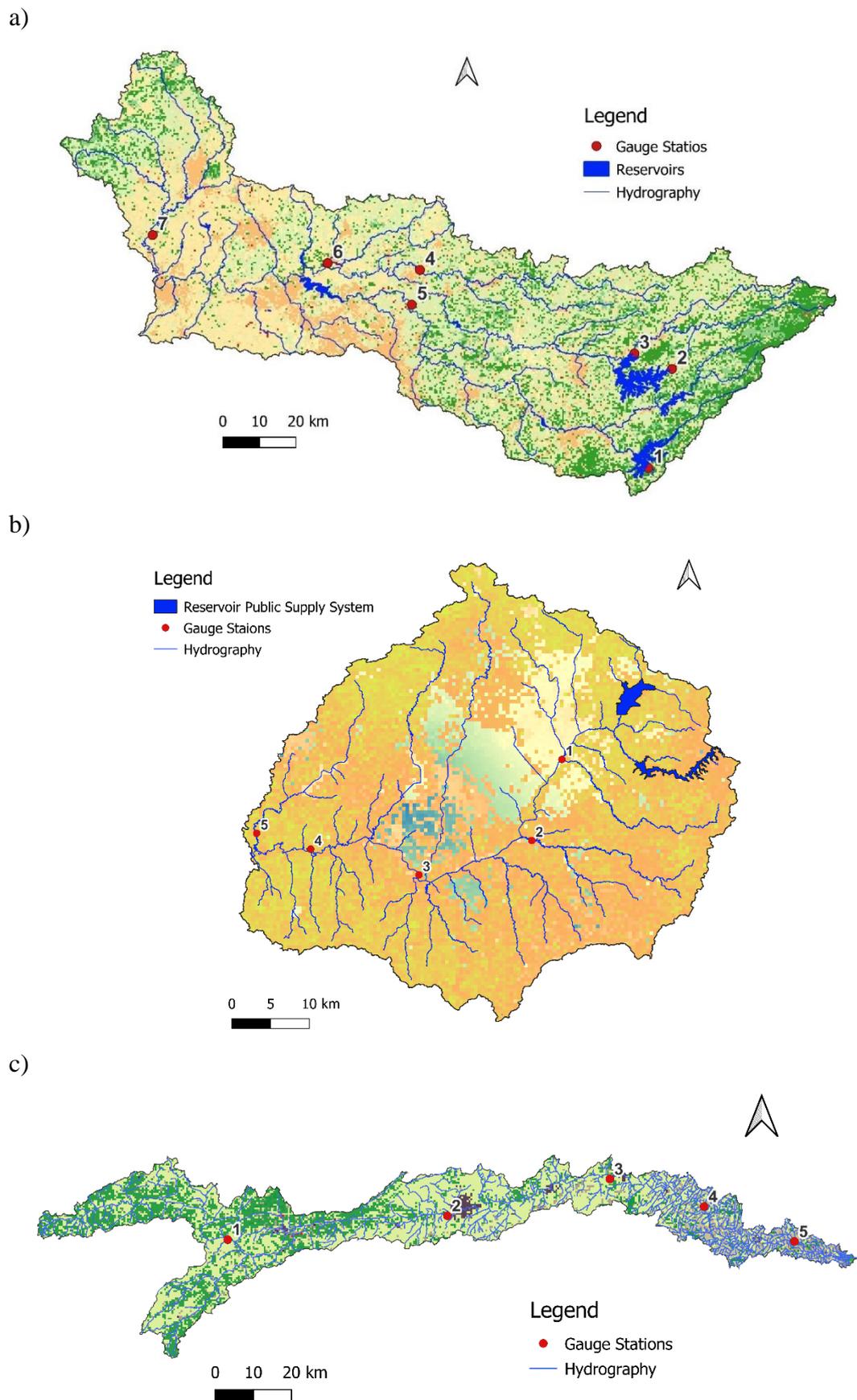
<b>Performance classification</b>	<b>RMSE</b>	<b>NSE</b>	<b>RB</b>
Very good	$\geq \frac{SD}{3.2}$	$\geq 0.9$	$\geq -0.12$ to $\leq 0.12$
Good	$\geq \frac{SD}{2.2}$ to $< \frac{SD}{3.2}$	0.8 to 0.9	$\geq -0.18$ to $< -0.12$ or $\geq 0.12$ to $< 0.18$
Acceptable	$\geq \frac{SD}{1.2}$ to $< \frac{SD}{2.2}$	0.65 to 0.8	$\geq -0.22$ to $< 0.18$ or $\geq 0.18$ to $< 0.22$
Unsatisfactory	$< \frac{SD}{1.2}$	$< 0.65$	$\leq -0.22$ or $\geq 0.22$

SD – standard deviation

## 5 MODEL APPLICATION RESULTS

This chapter describes the results of three selected basins for case study, obtained using RUBEM. PRB and UIRB are in the country's wet region of the southeast. Half of the IRB is in the dry region, and the other half is in the wet region in the country's northeast.

The calibration period was set between January-2000 to December-2009 and the validation period was January-2010 to December-2018. Figure S 7 presents the gauges stations used at each basin for calibration purpose. The streamflow information is scarce in these areas, however, data series with the less failures were selected. Additionally, the presence of large reservoirs at PRB and UIRB impacts the observed data.



**Figure S7.** Gauge Stations selected for calibration process and reservoirs at a) PRB, b) UIRB, c) IRB.

The reservoirs located close to gauge stations 1, 2 and 3 (Figure S 7a) in PRB basin, are a part of a complex water system that supply São Paulo city (the Cantareira system). A considerable amount of water is transposed from PRB to another basin, following an operation rule that consider the level of the reservoirs, by the rule, up to  $33 \text{ m}^3 \text{ s}^{-1}$  can be abstracted from PRB (ST 5) to Upper Tiete Basin.

The basins have different levels for water use, with the priority for public supplies, industry, and agriculture. Table S6 shows the average flow of the basins and the principal demands.

**Table S5.** Operation Rule for Cantareira system [88].

<b>Operation Range</b>	<b>Volume (V) of reservoirs system (accumulated)</b>	<b>Maximum flow abstracted (<math>\text{m}^3 \text{ s}^{-1}</math>)</b>
1	$V \geq 60\%$	33
2	$60\% > V \geq 40\%$	31
3	$40\% > V \geq 30\%$	27
4	$30\% > V \geq 20\%$	23
5	$V < 40\%$	15.5

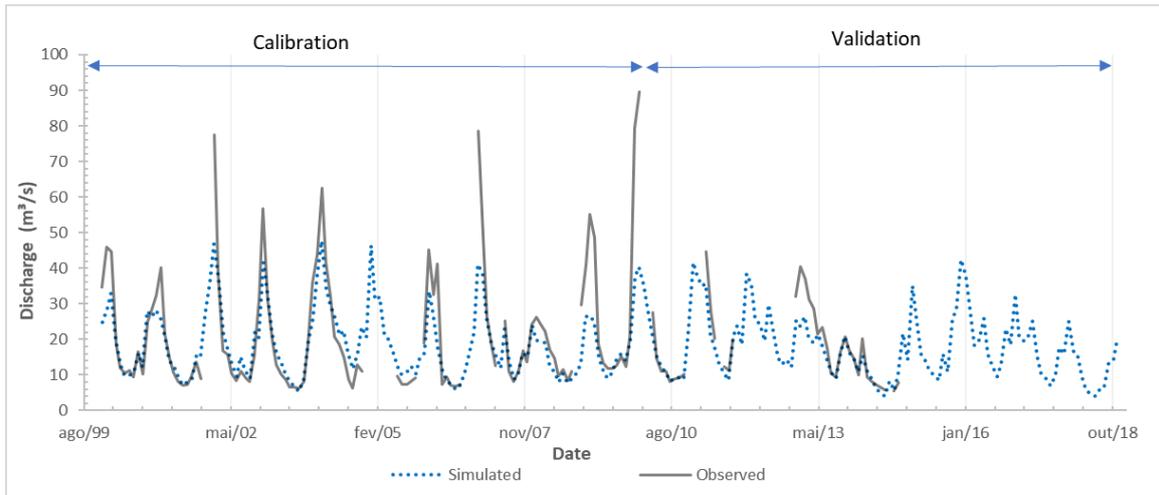
**Table S6.** Flow of the main water demands and discharge ( $\text{m}^3 \text{ s}^{-1}$ ) in the studied basins.

<b>Demand</b>	<b>Ipojuca River Basin<sup>3</sup></b>	<b>Piracicaba River Basin<sup>4</sup></b>	<b>Upper Iguaçú River Basin<sup>5</sup></b>
Public supply <sup>1</sup>	1.76	12.65	9.06
Industry	0.43	9.37	1.67
Agriculture <sup>2</sup>	1.73	6.74	0.25
Total	3.92	28.76	10.98
Average discharge	9.31	132.55	58.15

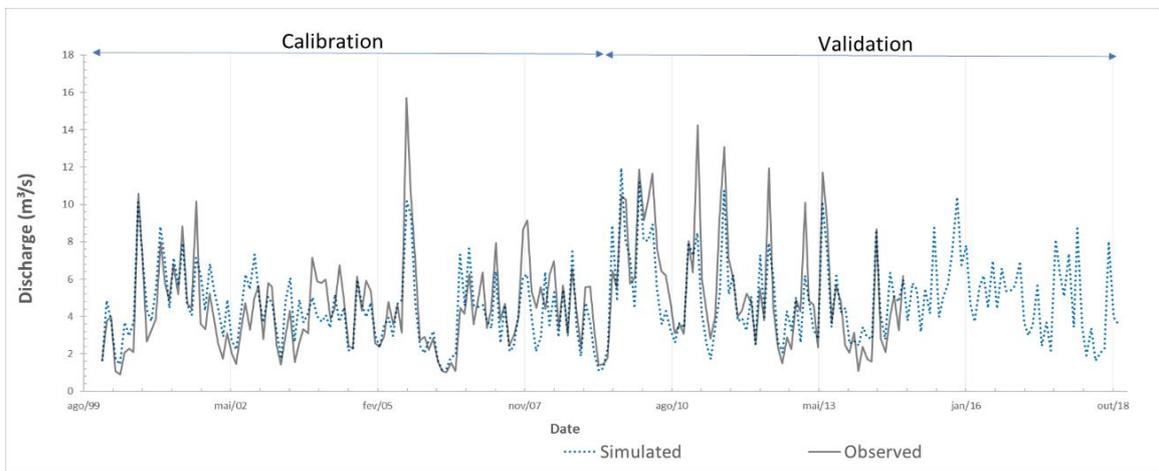
1 - Urban and rural; 2 - Irrigation and livestock; 3 –[127]; 4 – [88]; 5 – [89]

Figure S8 presents the hydrographs for gauge station 7 at PRB and 5 at UIRB. These stations are located at upper basins with less interference of water abstractions and absence of reservoirs. The NSE obtained in this case were 0.657 and 0.713 for calibration and validation at PRB gauge station, and 0.664 and 0.716 for calibration and validation at UIRB gauge station.

a)



b)



**Figure S8.** Hydrographs upper basins gauge stations: a) Point 7 in Piracicaba River Basin, and b) Point 5 in Upper Iguazu River Basin.

The parameters obtained during calibration were applied during the validation period. Table S7 shows the resulting parameters values for obtained for each study area.

**Table S7.** Parameter values of the RUBEM model obtained during calibration and applied to validation in the three studied basins.

<b>Parameter</b>	<b>IRB</b>	<b>PRB</b>	<b>UIRB</b>
Interception parameter ( $\alpha$ )	4.415	1.049	9.771
Parameter related to soil moisture ( $\beta$ )	0.078	0.152	0.181
Land use factor weight ( $w_1$ )	0.510	0.470	0.460
Soil factor weight ( $w_2$ )	0.120	0.350	0.430
Slope factor weight ( $w_3$ )	0.370	0.180	0.110
Regional Consecutive Dryness level ( $R_{CD}$ )	5.375	7.957	8.342
Flow direction factor ( $f$ )	0.581	0.767	0.831
Baseflow recession coefficient ( $\alpha_{gw}$ )	0.922	0.782	0.552
Flow recession coefficient ( $x$ )	0.307	0.219	0.107

Simulated discharges with parameters set were obtained in each basin at gauge stations location. In general, RMSE and NSE were classified as “accepted” most gauge station’s location in all study regions for both calibration and validation periods [62,56,59,71]. The RB was “unsatisfactory” in the IRB, which presented higher values in the calibration. In PRB, bias results were equal to or greater than acceptable at three-gauge stations in calibration period and four validation stations. In the UIRB, “good” or “very good” levels were observed during both periods, with a slight negative bias prevailing.

Table S8 presents efficiency indicators for downstream gauge stations of each basin. According to the criteria, RMSE was acceptable or good in both periods, for all the basins. NSE was “acceptable” for UIRB and “unsatisfactory” for IRB and PRB in the calibration period, according to [71]. Two out of five-gauge stations used for IRB calibration, are in the semiarid region, and three in the wetland. According to [72], global calibration is satisfactory when limited to homogeneous hydrological regimes, otherwise can result in a poor calibration performance, specifically for estimating high flows in small catchments. Four out of seven-gauge stations selected for PRB calibration, are located downstream from the regularization reservoirs, where water is transferred to another basin. These factors contributed to reducing the NSE in the calibration period. In the validation period, the NSE was “acceptable” or “good” in the three basins. RB was “very good” for UIRB during calibration and IRB during validation. The discharges were overestimated in PRB and underestimated in IRB and UIRB.

The observed e simulated discharge were compared using asynchronous regression (Figure 4 on manuscript) of the accumulated distribution percentiles [60], the results show  $R^2 > 0.89$  for all basins, with the tendency of the model to underestimate flows in the IRB and UIRB basins and to underestimate in the PRB. Figure S10 shows adjustments obtained in the studied basins. The fitted line is only effective for comparison with the 1:1 ratio between observed and simulated values. [64] used linear correlation to assess SWAT calibration and validation in a Brazilian basin, with values considered satisfactory for  $R^2 = 0.72$ . We performed the Fisher's test of the two samples considering the indifference of the mean and variance

between the series as the null hypothesis. It was found that there is no significant difference (99% confidence) between the observed and simulated flows in the three basins.

**Table S8.** Efficiency indices of the RUBEM in drainage areas of gauge stations of the studied basins during calibration and validation periods.

Basin	Area (km <sup>2</sup> )	Calibration					Validation				
		N	SD	RMSE	NSE	RB	N	SD	RMSE	NSE	RB
IRB	3310	82	16.144	<u>12.32</u>	0.417	-0.397	105	9,070	<u>5.355</u>	<u>0.652</u>	<b>0.007</b>
PRB	3400	120	23.524	<u>17.470</u>	0.448	0.468	60	39,926	<u>21.303</u>	<u>0.715</u>	0.348
UIRB	2330	120	26,250	<u>13.37</u>	<u>0.741</u>	<b>-0.027</b>	108	22,702	<b>8.072</b>	<b>0.874</b>	<b>-0.14</b>

Legend: black number = Unsatisfactory; black underlined number = Acceptable; bold number = Good; bold underlined number = Very good; N - sample size; SD – standard deviation.

The observed and simulated mean discharge during the dry and wet season in the three basins for the entire analysis period (January 2000 to December 2018) are shown in Table S9. It is observed that RUBEM underestimated the mean discharges by 23.5% in wet seasons and by 9.3% in dry seasons for IRB. In PRB, mean discharge was overestimated by 33.7% and 51%, respectively, in wet and dry seasons. In UIRB, the model underestimated the mean discharge by 8.1% in wet seasons and overestimated by 4.3% in dry seasons.

**Table S9.** Mean discharges in dry and wet seasons ( $\text{m}^3 \text{s}^{-1}$ ).

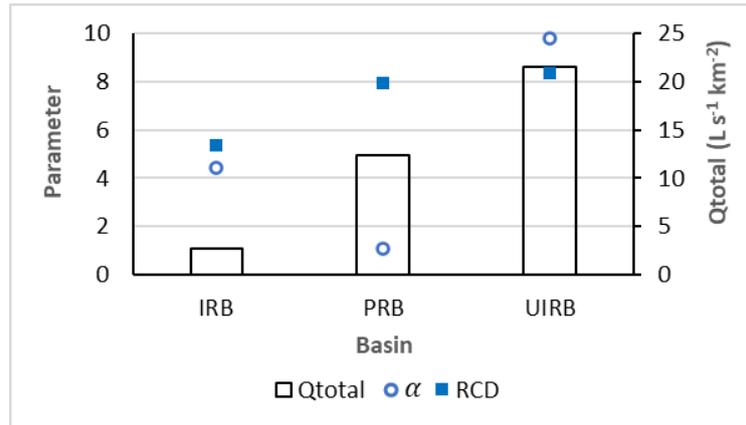
Season	IRB		PRB		UIRB	
	Observed	Simulated	Observed	Simulated	Observed	Simulated
Dry	7.1	6.4	16.6	34.0	42.1	44.0
Wet	14.8	11.3	36.3	54.8	59.2	54.5

The specific mean discharge simulated at the downstream gauge station in the basins and the calibrated parameters are shown in Figure S10. The basins located in the wet region (PRB and UIRB) have a greater specific discharge than those in the dry region (IRB). The interception parameter for UIRB ( $\alpha = 9.77$ ) was significantly higher, indicating a greater water storage capacity in the vegetation canopy, resulting in higher interception of rainfall. The continuous level of drought ( $RCD$ ) and the weight factors for soil land use ( $w1$ ), moisture ( $w2$ ), and slope ( $w3$ ) are related to the resulting surface runoff. The values of  $RCD$  and  $w2$  were higher in UIRB ( $RCD = 8.34$ ;  $w2 = 0.43$ ) and PRB ( $RCD = 7.96$ ;  $w2 = 0.35$ ). These parameters contribute to increase the potential runoff in permeable areas. In IRB basin, parameters  $RCD = 5.38$ ,  $w2 = 0.12$ , and low soil moisture factors ( $b = 0.08$ ) reduce surface runoff. The IRB's largest drainage area in the semiarid region explains the parameter's value.

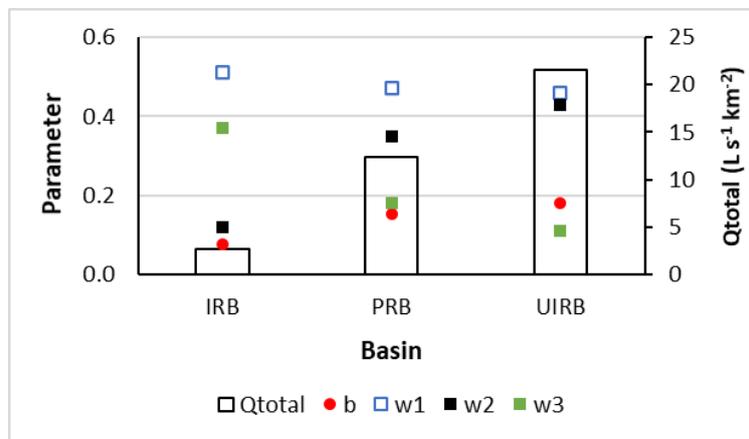
The high values for the partition of flow factor in PRB ( $f = 0.77$ ) and UIRB ( $f = 0.83$ ) favors the lateral flow over the recharge in these basins. The threshold base ( $B_{thresh}$ ) flow adopted in the three basins was 150 mm. The higher value of the decay coefficient in the IRB ( $\alpha_{gw} = 0.92$ ) decrease the base flow dependency on the previous month flow, and increase its dependency on the recharge, as expected in semiarid regions. In PRB and UIRB basins, the lowest values of this coefficient balance the influence of antecedent flow and recharge on base flow.

Higher values of the damping coefficient increase the importance of the previous month discharge and reduce the current month discharge generated in each cell. This coefficient resulted in higher value for IRB ( $x = 0.31$ ), causing more significant damping in discharge into the basin.

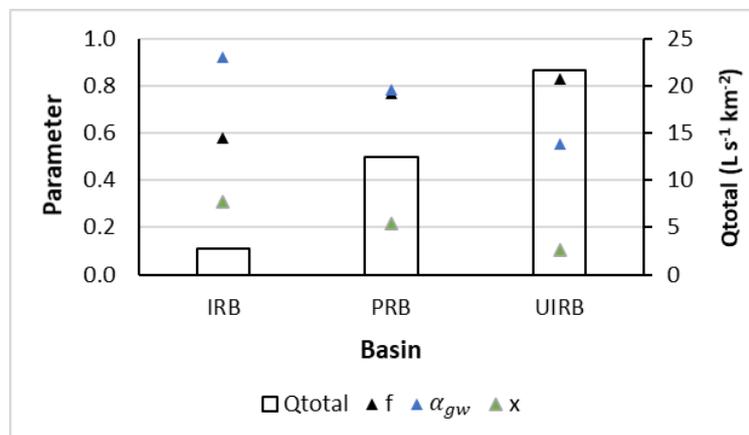
a)



b)



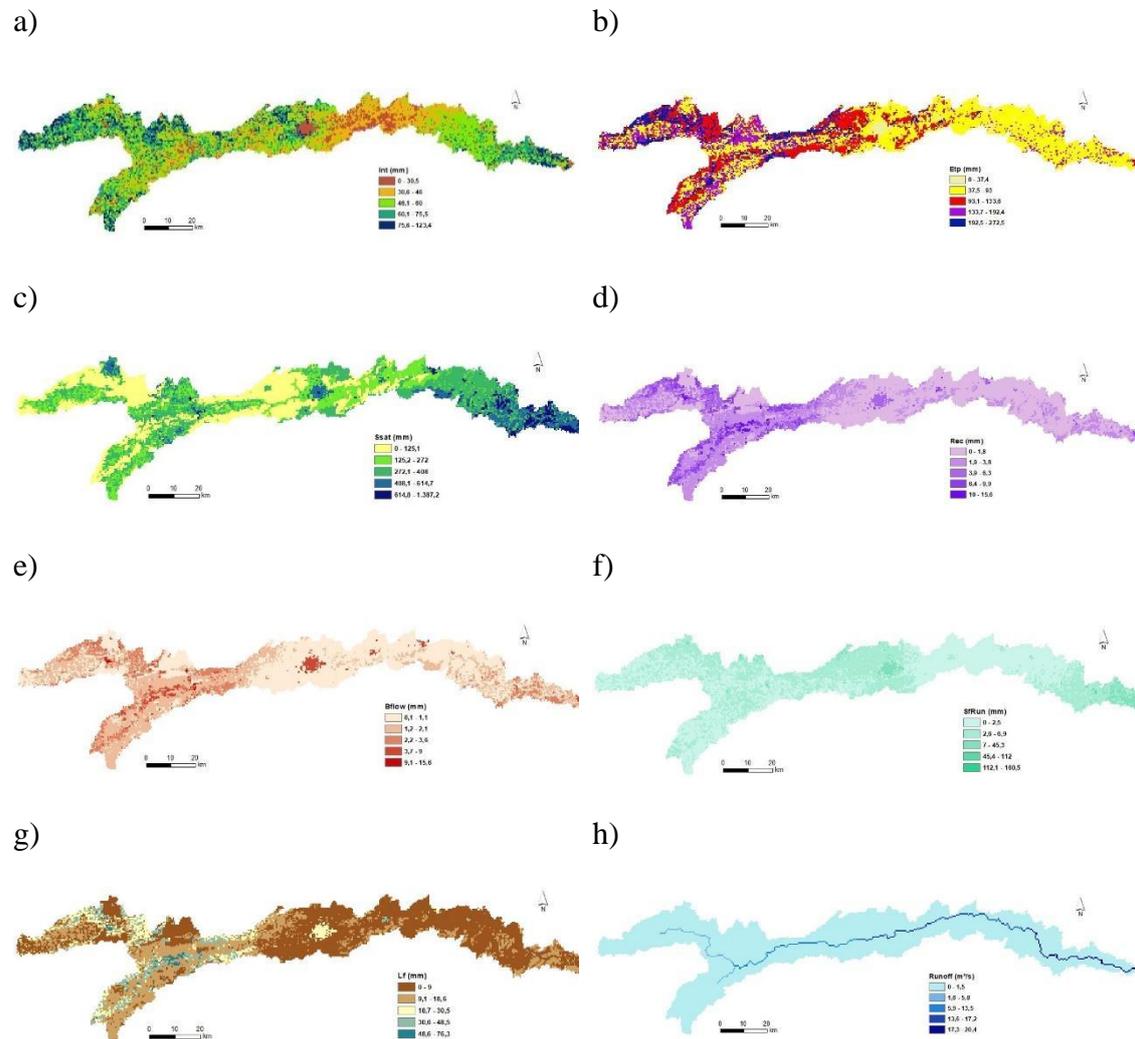
c)



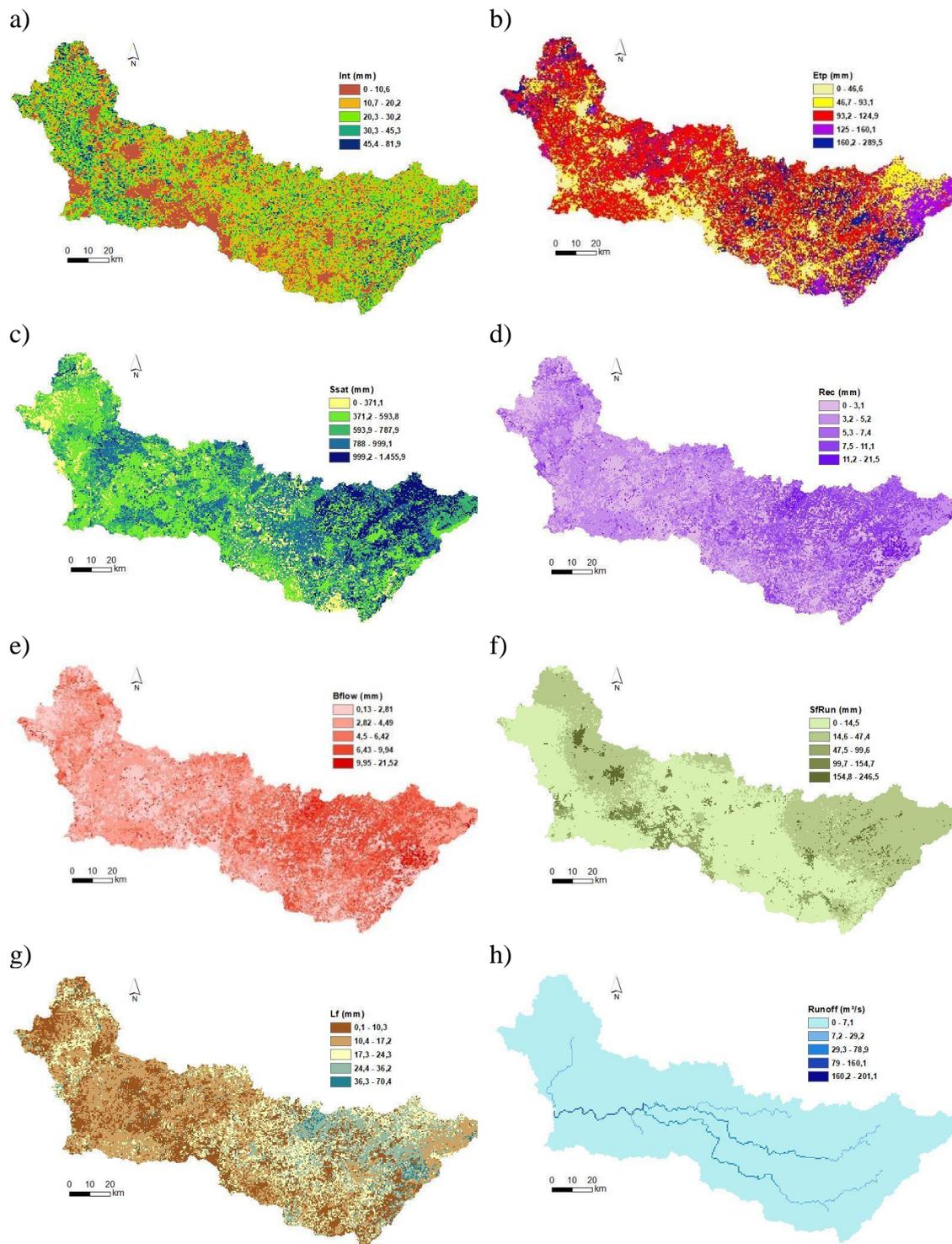
**Figure S10.** Total specific mean discharge and calibrated parameters in the basins: a) interception ( $\alpha$ ) and regional consecutive dryness factor (*RCD*); b) soil moisture (*b*), weights for the components related to runoff: cover (*w1*), soil (*w2*) and slope (*w3*); c) vertical and horizontal flow partitioning coefficient (*f*), base flow decay coefficient ( $\alpha_{gw}$ ) and damping coefficient (*x*).

RUBEM outputs are spatial raster file format results. In addition, graphical results at any user indicated location is available. The temporal series of the resulting hydrological variables are provided in raster format. The spatialized results of variables

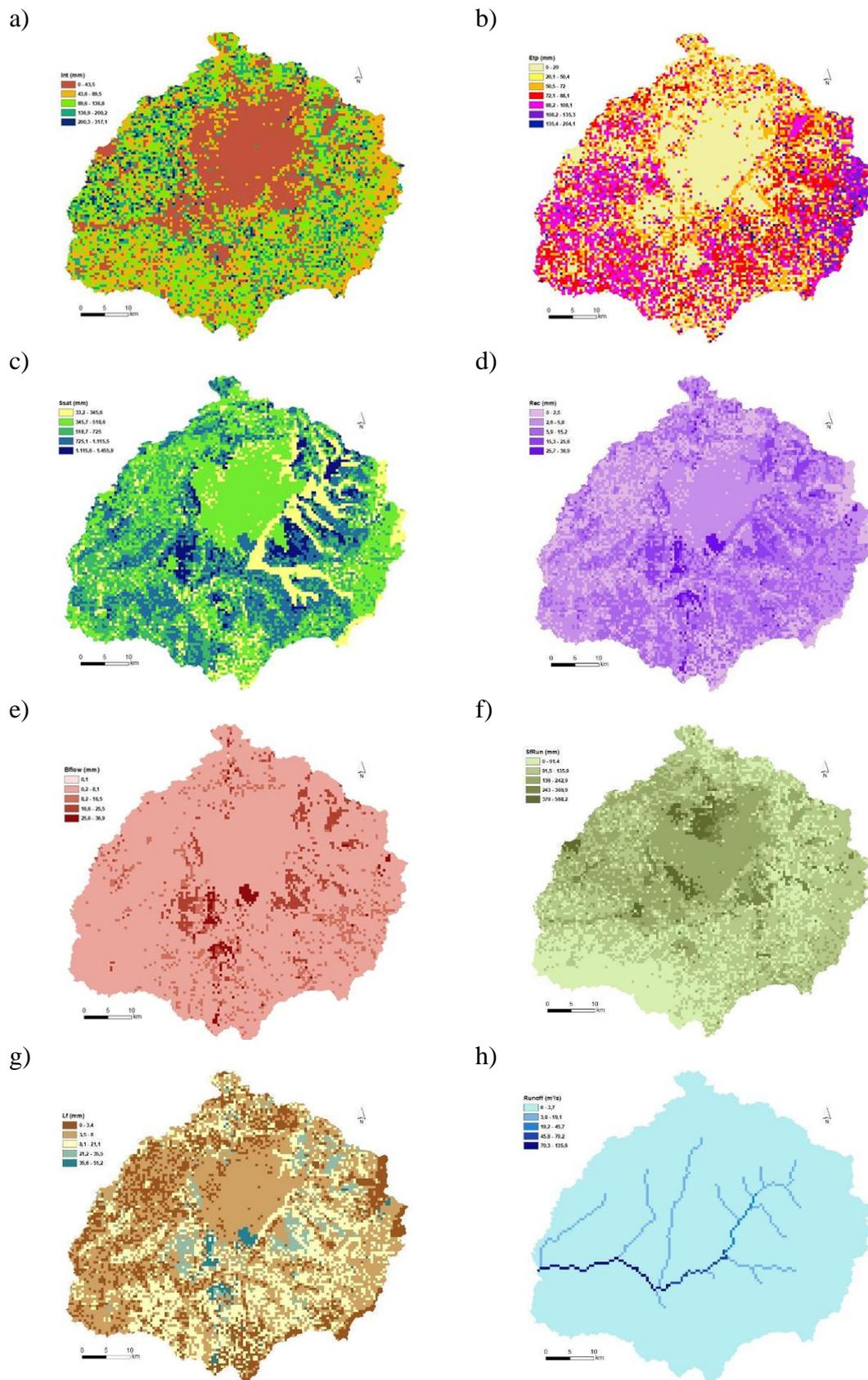
calculated by the model are presented for IRB (Figure S11), PRB (Figure S12), and UIRB (Figure S13).



**Figure S11.** RUBEM model outputs for IRB in April 2008: a) vegetation cover interception, b) evapotranspiration, c) root zone moisture content, d) recharge, e) base flow, f) surface runoff, g) lateral flow, and h) total discharge.



**Figure S12.** RUBEM model outputs for PRB in March 2009: a) vegetation cover interception, b) evapotranspiration, c) root zone moisture content, d) recharge, e) base flow, f) surface runoff, g) lateral flow, and h) total discharge.



**Figure S13.** RUBEM model output for UIRB in January 2010: a) vegetation cover interception, b) evapotranspiration, c) root zone moisture content, d) recharge, e) base flow, f) surface runoff, g) lateral flow, and h) total discharge.

## 6 PLUGIN MODEL INSTALLATION

RUBEM Hydrological was originally developed into QGIS 3.x plug-in, it is strongly recommended to use the latest LTR version of QGIS 3.x. The latest version of the plug-in is available on the GitHub repository release page (<https://github.com/LabSid-USP/RUBEMHydrological/releases>) or directly from the QGIS Python Plugins Repository<sup>7</sup>.

A video tutorial procedure of RUBEM Hydrological plug-in installation is available on LabSid YouTube channel (<https://youtu.be/F8zx9So2nrI>). In addition, the plug-in's online documentation provides installation and usage information (<http://rubem-hydrological.readthedocs.io/>).

---

<sup>7</sup> The submission of the RUBEM Hydrological plug-in to the QGIS Python Plugins Repository is under evaluation by the development team.