

# **Tropical woodland moisture cycling responses to climate stochasticity using eddy covariance and machine learning**

**Jamil A.A. Anache<sup>1,2</sup>, Livia M.P. Rosalem<sup>1</sup>, André S. Ballarin<sup>1</sup>, Alex N.A. Kobayashi<sup>1</sup>,  
Alessandra C. Santos<sup>1</sup>, and Edson Wendland<sup>1</sup>**

<sup>1</sup> São Carlos School of Engineering (EESC), University of São Paulo (USP), São Carlos, SP, 13566-590, Brazil.

<sup>2</sup> Federal University of Mato Grosso do Sul (UFMS), CxP. 549, Campo Grande, MS, 79070-900, Brazil.

Corresponding author: Jamil A.A. Anache (jamil.anache@ufms.br)

## **Key Points:**

- Water fluxes of a tropical woodland measured using eddy covariance are reproduced using machine learning for long-term simulations.
- The soil water in the wooded Cerrado helps to maintain the moisture cycling over the dry season for the ecosystem and surroundings.
- The probability of a dry year in the wooded Cerrado is close to 20%, revealing a return period of approximately 5 years to droughts.

**Abstract**

Brazilian tropical woodlands, such as the wooded Cerrado, perform hydrological functions that need to be well understood by field data acquisition and mathematical modeling. Here, we aimed to assess the water partitioning behavior and variability of a wooded Cerrado fragment located in Southeastern Brazil by (i) measuring fluxes using eddy covariance; (ii) applying machine learning techniques to obtain a model to estimate the evapotranspiration (ET) using meteorological data as input: solar radiation ( $R_g$ ), wind speed (WS), temperature (T), relative humidity (RH), and rainfall (P); and (iii) simulating a long-term water balance using stochastic climate generator inputs and the previously calibrated ET model. The average observed ET was  $3.12 \pm 0.93 \text{ mm d}^{-1}$ , and the uncertainties were close to  $\pm 20\%$ . We found that the k-nearest neighbors (KNN) presented the best performance to obtain a ET model, when compared to other machine learning techniques: decision trees (DTR), neural network multilayer perceptron (MLP), and ensemble based AdaBoost (ADA). The simulated long-term water balance at a daily basis was more sensitive to variations in P than ET and presented a resilient behavior to droughts, showing the importance of tropical woodlands to maintain the moisture cycling throughout the year with the water stored in the subsurface. Finally, the water balance in the long-term relies on the rainfall stochasticity, as the average annual  $dS/dt$  (water balance residual) ( $172 \pm 211 \text{ mm yr}^{-1}$ ) and P ( $1227 \pm 208 \text{ mm yr}^{-1}$ ) have similar standard deviations; differently from the ET ( $1054 \pm 46 \text{ mm yr}^{-1}$ ), which presented higher annual rates with a small variability throughout the simulation.

**Keywords:** Water balance, wooded Cerrado, regression tree, neural networks, K-nearest neighbors, AdaBoost, weather generator.

## 1 Introduction

Globally, evapotranspiration is commonly estimated using mass conservation approaches, however the distribution of this variable along the Earth's surface still needs site specific monitoring and modeling (Wang & Dickinson, 2012). Eddy covariance instrumentation and processing techniques are the actual backbone of evapotranspiration (ET) measurements worldwide, as a reference to provide more reliable hydrological models in multiple scales and principles: remote sensed (Bastiaanssen et al., 1998; Martens et al., 2017; McCabe & Wood, 2006; Mu, Zhao, & Running, 2011; Talsma et al., 2018); and ground based (Allen, Pereira, Raes, & Smith, 1998; Priestley & Taylor, 1972). Eddy covariance instruments generate complex flux measurements that can be better reproduced using new advanced machine learning techniques (Baldocchi, 2020), outperforming traditional evapotranspiration estimation methods (Goyal, Bharti, Quilty, Adamowski, & Pandey, 2014). Thus, machine learning has become a popular approach to estimate hydrological or environmental variables, enhancing data processing and interpretation (Yamaç & Todorovic, 2020).

Brazil is recognized as a huge biodiversity hotspot, as its biomes contain resources that are useful for many economic activities, and regulate global climate, supplying Earth with a number of ecosystem services (Levis et al., 2020). Regarding the landscapes that form the Brazilian Cerrado, a better understanding of evapotranspiration is necessary to improve estimation techniques, which would be useful for developing unbiased hydrological models for more effective water resources management (Valle Júnior et al., 2020), in order to sustain economic activities, such as agriculture and livestock production.

Many efforts were made for a better understanding of the water cycle in Brazil on multiple scales (Melo et al., 2020; P. T. S. Oliveira et al., 2014). These studies contain valuable water partitioning data from tropical regions, but this type of information is still scarce for these regions (Burt & McDonnell, 2015; Wohl et al., 2012). Nevertheless, evapotranspiration can be a significant source of uncertainty, presenting more than 50% of bias among the estimates and observations (Anache, Wendland, Rosalem, Youlton, & Oliveira, 2019). The Large-Scale Biosphere-Atmosphere Experiment in the Amazon (da Rocha et al., 2009) and individual efforts to measure evapotranspiration in undisturbed Cerrado areas have been carried out recently (Cabral, da Rocha, Gash, Freitas, & Ligo, 2015; Giambelluca et al., 2009; Nobrega et al., 2017). These studies revealed some important results that led to the validation of many

evapotranspiration estimates in the region (Oliveira et al., 2015; A. L. Ruhoff et al., 2013; Anderson L. Ruhoff et al., 2012).

Considering the Twenty-three unsolved problems in hydrology (UPH) (Blöschl et al., 2019), an important question arises in this study: “How can hydrological models be adapted to be able to extrapolate changing conditions, including changing vegetation dynamics?” In order to solve this UPH, weather generators are commonly used to assess the impact of climate variability on previously calibrated hydrological models (King, McLeod, & Simonovic, 2015; Kwon, Sivakumar, Moon, & Kim, 2010). Synthetic data generated by these models provide significant information to decision makers in places with a coarse spatiotemporal resolution of the observed data. These stochastic tools can reinforce the available data as they can fill the gaps of missing data or provide long-term series with the same characteristics observed in the input variables, such as variances, correlations and spatiotemporal dynamics (Allard, Ailliot, Monbet, & Naveau, 2015; King et al., 2015). Moreover, weather generators can simulate extreme conditions expected in the location but not observed in the historical series, providing a wider range of feasible situations (Herrera et al., 2017; Semenov, Brooks, Barrow, & Richardson, 1998).

These stochastic tools can be classified into parametric, semi-parametric and non-parametric ones. Traditional models, also known as parametric models, require assumptions about statistical distributions of the input weather variables, involving a careful analysis of the suitability of the output data to the location (Herrera et al., 2017). Some of parametric weather generators widely used are the CLIGEN model (Nicks, Lane, & Gander, 1995) and the WGEN (Richardson & Wright, 1984). Non-parametric weather generators do not require any statistical assumption of the observed data, creating climate simulations based on resampling methods. Semi-parametric ones, such as LARS-WG (Semenov & Stratonovitch, 2010), were developed considering both parametric and non-parametric components, designed to overcome possible limitations of using parametric generators (Apipattanavis, Podestá, Rajagopalan, & Katz, 2007).

There is a great need to reduce uncertainties of evapotranspiration observations and estimates performed for tropical woodlands such as the Brazilian wooded Cerrado (Anache et al., 2019). Additionally, few studies in Brazil and in South America have endeavored to represent a hydrological phenomenon such as evapotranspiration using machine learning techniques

(Tyralis, Papacharalampous, & Langousis, 2019). Finally, the extrapolation of climate conditions should be more included in hydrological models that can catch the behavior of undisturbed ecosystems in the outcomes. Thus, the purpose of this study was to assess the water partitioning variability in the critical zone of a wooded Cerrado fragment due to climate stochasticity. Firstly, we performed flux measurements over a wooded Cerrado area located in Southeastern Brazil using eddy covariance; secondly, we explored the data using machine learning to obtain a model to calculate the evapotranspiration of the wooded Cerrado area using meteorological data feasible to be estimated or measured; finally, we simulated a long-term water balance using stochastic climate generator inputs and the previously calibrated model.

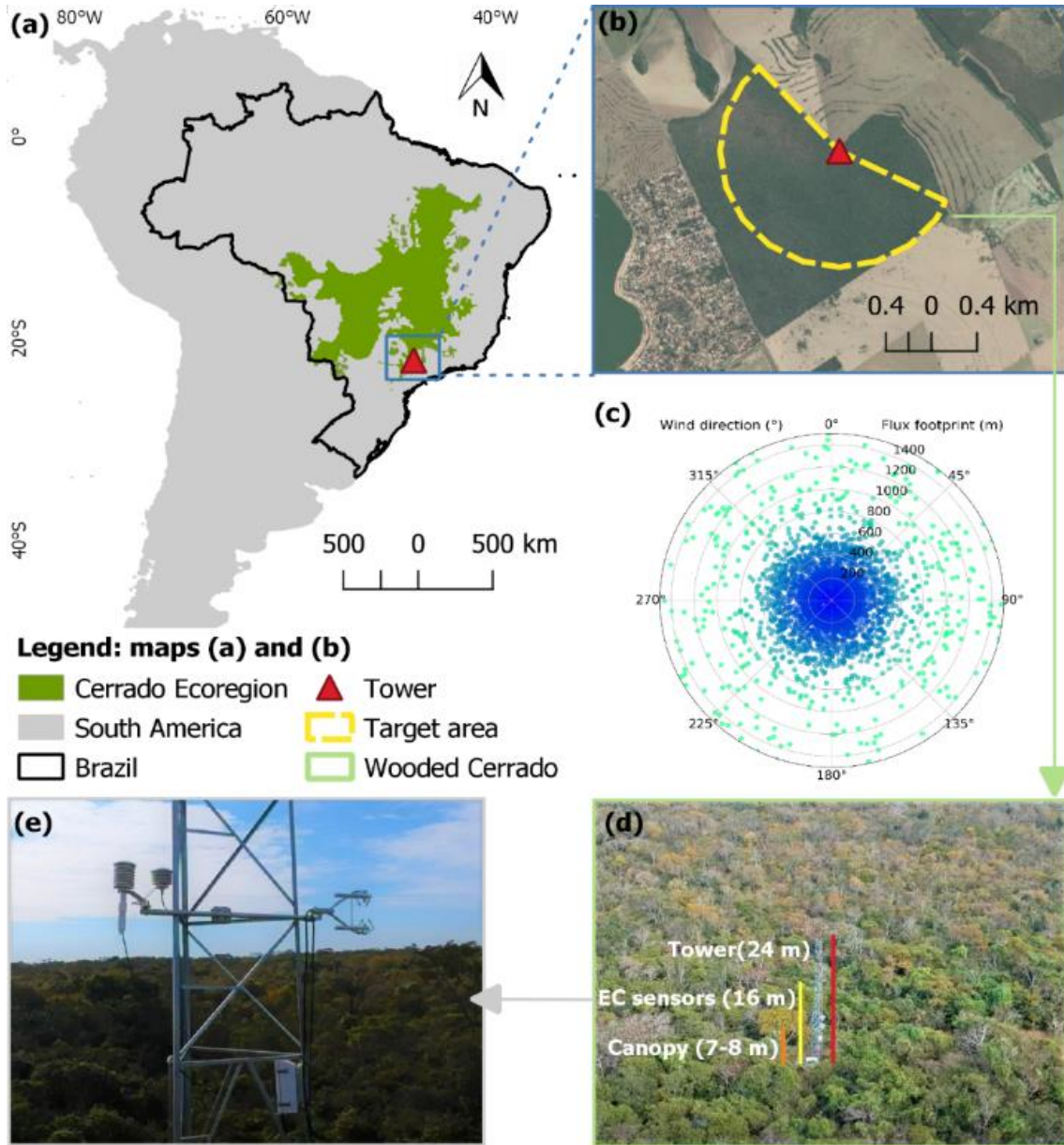
## 2 Material and Methods

### 2.1 Site description

The experimental site (Figure 1) was built inside a tropical woodland fragment of approximately 330 ha commonly known as wooded Cerrado (or Cerrado *stricto sensu*, in Portuguese). The area is in the central region of the state of São Paulo, in Brazil, where pastureland and sugarcane fields are the most common land covers in the surroundings. The site is located inside the Cerrado ecoregion, which is part of Tropical and Subtropical Grasslands, Savannas and Shrubland biomes (Olson et al., 2001). The average rainfall is 1486 mm (Cabrera, Anache, Youlton, & Wendland, 2016) and the climate is humid subtropical (Cwa) according to Köppen's classification system (Peel, Finlayson, & McMahon, 2007).

### 2.2 Instrumentation and monitoring

Micrometeorological variables were monitored using low (rainfall, relative humidity, temperature, net radiation, soil heat flux, soil moisture) and high (wind speed and direction, H<sub>2</sub>O molar fraction) frequency instrumentation. The instruments (Table 1) were deployed in a 24-meter height metallic tower, which was equipped with a data acquisition and storage system. However, the main pieces of equipment were positioned below the maximum tower height due to footprint requirements and lightning protection.



**Figure 1.** Site regional context (a) and location (b); wind direction and flux footprint (c); canopy, instruments, and tower height (c); and eddy covariance sensors above canopy (e).

**Table 1.** Monitoring site instrumentation features.

Variable	Sensor	Height or depth* (m)	Measurement range	Maximum error
Temperature (°C)	HMP155A	16	-80 to +60 °C	±0.45 °C
Relative humidity (%)	HMP155A	16	0 to 100 %	±1.7 %
Rainfall (mm)	Hydrological Services TB4	16	0 to 700 mm h <sup>-1</sup>	±3 %
Atmospheric pressure (mbar)	Vaisala CS106	16	500 to 1100 mbar	±1.5 mbar
Wind speed (m s <sup>-1</sup> ) and direction (°)	Irgason sonic anemometer	16	0 to 30 m s <sup>-1</sup> 0 to 360°	±1.8 m s <sup>-1</sup> ±0.7°
H <sub>2</sub> O molar fraction	Irgason gas analyzer	16	0 to 72 mmol/mol	2%
Soil moisture (%)	FDR EnviroSCAN Sentek	-0.3*	0 to ~65 %	±3 %
Net solar radiation (W m <sup>-2</sup> )	Kipp & Zonen CNR4	10	±2000 W m <sup>-2</sup>	±20 W m <sup>-2</sup>
Soil heat flux (W m <sup>-2</sup> )	Hukseflux HFP01	-0.1*	±2000 W m <sup>-2</sup>	-15 % to +5 %

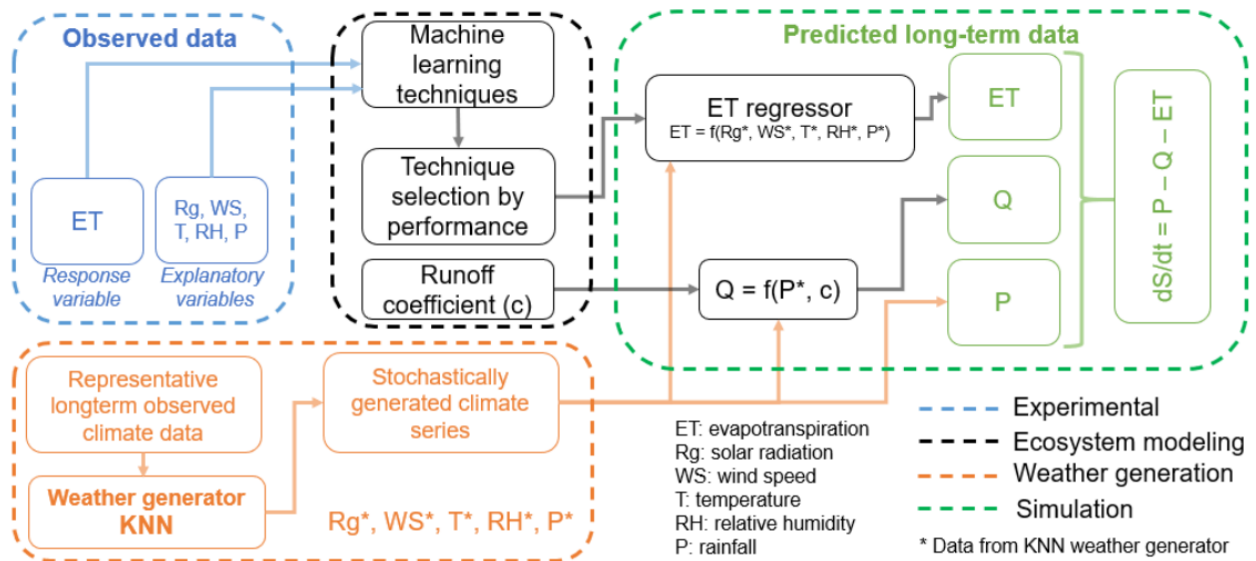
### 2.3 Observed data processing and analysis

The raw data of observed evapotranspiration (latent heat flux) and sensitive heat were corrected by using EasyFlux-PC (*Campbell Scientific Inc.*). The main corrections steps were: (i) peak removals and filtering throughout the time series, considering signal strength and equipment detection limits; (ii) correction of the horizontal winds using the double rotation method (Tanner & Thurtell, 1969); (iii) lag removal between concentration and wind measurements (Foken, Leuning, Oncley, Mauder, & Aubinet, 2012; Horst & Lenschow, 2009); (iv) averaging and frequency corrections (Dijk, 2002; Foken et al., 2012; Horst & Lenschow, 2009; Kaimal, Clifford, & Lataitis, 1989; Moncrieff et al., 1997; Montgomery, 1947; Moore, 1986; Shapland, Snyder, Paw U, & McElrone, 2014); (v) conversion of the sonic sensible heat flux to sensible heat flux (Dijk, 2002; Schotanus, Nieuwstadt, & De Bruin, 1983); (vi) air density fluctuation correction (Webb, Pearman, & Leuning, 1980); (vii) data quality control (Foken et al., 2012).

Finally, the fluxes were filtered using the wind direction as the main criteria to select measurements that represent the target area (a tropical woodland). Fluxes that came from directions outside the woodland fragment or without the required fetch were considered as gaps throughout the series. Thus, we filled the gaps using the ReddyProc webtool developed by Max-Planck Institute for Biogeochemistry (Wutzler et al., 2018) to obtain a consistent time series.

## 2.4 Modeling phase

Observed micrometeorological data collected through the eddy covariance approach was used as input data to calibrate and validate several modeling strategies using machine learning to model evapotranspiration. After choosing the best modeling strategy, we performed a long-term water balance simulation using a weather generator to provide a huge variety of input data to run the previously calibrated ET model, and also calculated overland flow with the simulated rainfall data (Figure 2). The evapotranspiration is a key variable in the water balance to represent the hydrological response of an ecosystem, as well as the overland flow, which can be easily estimated by using a literature runoff coefficient.



**Figure 2.** Workflow chart containing the modeling phases of the study. 2.4.1 Machine learning for ET estimation.

Half-hourly data obtained after processing the high frequency eddy covariance measurements were downscaled to daily data (Anache et al., 2021). We defined as the evapotranspiration (ET) the target (response) variable, which can be explained by selected explanatory variables also monitored in the same site: incoming solar radiation (Rg), wind speed (WS), temperature (T), relative humidity (RH), and rainfall (P). In addition, these variables were ranked according to their importance in estimating ET using a recursive feature elimination algorithm. The whole dataset, which has 235 days in total was randomly and equally divided into training (calibration) and testing (validation) datasets. We included all seasonal conditions in



both training and testing datasets: 117 observed points (days) were used for training, and the rest (118) were used for testing. The subsets were created randomly and interspersed throughout the 235 days of the data collection period.

We tested four machine learning techniques available in the scikit-learn machine learning open-source library (Pedregosa et al., 2011), which works using Python. We also adopted unsupervised learning to fit the tested models. The tested learning techniques were regressions based on decision trees (DTR), neural network multilayer perceptron (MLP), k-nearest neighbors (KNN), and ensemble based AdaBoost (ADA). The machine learning techniques performances were evaluated by using statistical metrics: coefficient of determination ( $R^2$ ), mean squared error (MSE), percent bias (PBIAS), and Kling-Gupta efficiency (KGE).

#### 2.4.2 Overland flow estimation

We used a runoff coefficient (0.001) previously calculated for the same location and land cover (Anache et al., 2019) to simulate the overland flow generation. A previous study in the same site recommends the rational method as the most feasible technique to estimate the overland flow (P. T. S. Oliveira et al., 2016) due to the higher uncertainties of other methods. Additionally, in forested conditions, the overland flow is not significantly as high as the other water balance components (Anache et al., 2019; Anache, Wendland, Oliveira, Flanagan, & Nearing, 2017; Guo, Hao, & Liu, 2015; Nacinovic, Mahler, & Avelar, 2014).

#### 2.4.3 Weather generator for long-term water balance simulation

We used the KnnCADv4 model (King et al., 2015) to generate a synthetic daily series of rainfall, maximum, average, and minimum temperature, solar radiation, wind speed, and relative humidity. These variables are important inputs for ET models. This weather generator is typically used to simulate temperature and precipitation data considering climate stochasticity. These outcomes can be used as input parameters to run previously calibrated models to obtain other specific variables and transfer the variability from the input series (Herrera et al., 2017). We used the climate dataset from the Luiz de Queiroz College of Agriculture (ESALQ-USP) climate station in Piracicaba, Brazil, established in 1917 (latitude 22°42'30" S, longitude 47°38'30"W, elevation 546 m a.m.s.l.) (Marin, Pilau, & Sentelhas, 2021) to set up the KnnCADv4 model and obtain a long-term synthetic series. This station represents the regional

climate of the study site due to its long-term characteristic and location (Anache, Flanagan, Srivastava, & Wendland, 2018).

The modified algorithm of the KnnCADv4 model (King et al., 2015; Mandal, Arunkumar, Breach, & Simonovic, 2019) is a non-parametric weather generator that uses a resampling strategy for generating data based on the K-Nearest Neighbor (K-NN) procedure. We opted for a non-parametric model due to its greater applicability than parametric and semi-parametric ones (Herrera et al., 2017). The KnnCADv4 generates simulated daily data based on the similarity with the historical data (Yazd, Salehnia, Kolsoumi, & Hoogenboom, 2019). Considering this, the model selects the possible nearest neighbors of a given day of the year, based on a temporal window with a predefined length, centered on that day for all years of the input record. These potential neighbors are then ranked, based on their Euclidean distance from the current day, and the first K observed data of them are selected. To preserve temporal correlation, resamples on KnnCADv4 are made based on a block with Y days as opposed to a single day (Mandal et al., 2019). The model also permits perturbing the resampled values, allowing values not observed in the input data in the generated series. More details about this tool can be seen in King et al. (2015) and Mandal et al. (2019).

The water balance variables were simulated using KnnCADv4 outcomes at a daily basis:  $R_g$ ,  $WS$ ,  $T$ ,  $RH$ , and  $P$  were the inputs to calculate  $ET$  using the machine learning technique that presented the best model performance metrics; the  $OF$  was calculated using simulated  $P$  and the wooded Cerrado runoff coefficient (0.001) (Anache et al., 2019); and the  $P$  used in the water balance simulation was the same as the KnnCADv4 outcomes. Thus, these variables were combined in Equation 1, to obtain the water balance residual ( $dS/dt$ ), which includes subsurface flow, soil water storage, deep percolation, and groundwater recharge.

$$\frac{dS}{dt} = P - ET - OF \quad (1)$$

Where:  $P$  is the rainfall ( $\text{mm d}^{-1}$ );  $ET$  is the evapotranspiration ( $\text{mm d}^{-1}$ );  $OF$  is the overland flow ( $\text{mm d}^{-1}$ ); and  $dS/dt$  is the water balance residual ( $\text{mm d}^{-1}$ ).

## 2.5 Data analysis and uncertainties

We tested the normality assumption for the daily basis datasets (observations and simulations) considering a 95% confidence interval. The observed data (P, ET, and soil water content) and observed derived data (OF and dS/dt) were plotted using a daily timescale. The soil water content (SWC) monitoring throughout the observational period was useful to validate the water balance assumption.

The uncertainties were assessed for the main observed hydrological outputs (P, ET, OF, and dS/dt). The flaws considered for P were derived from the instrument relative error (Table 1), and the OF error was calculated according to the relative error reported with the runoff coefficient study (Anache et al., 2019) and it is  $\pm 10\%$ . The error from ET measurements were calculated by EasyFlux-PC (*Campbell Scientific Inc.*) as random errors accumulated throughout the measurements and data processing. Hence, the water balance residual (dS/dt) uncertainty could be calculated as a propagation of errors (Equation 2). Then, the errors of the water balance components were accumulated throughout the period and expressed at a daily basis as dS/dt uncertainty.

$$\sigma_{\frac{dS}{dt}} = (\sigma_P^2 + \sigma_{ET}^2 + \sigma_{OF}^2)^{0.5} \quad (2)$$

Where:  $\sigma$  is the standard error ( $\text{mm d}^{-1}$ ); dS/dt is the water balance residual; P is rainfall; OF is surface runoff; and ET is evapotranspiration.

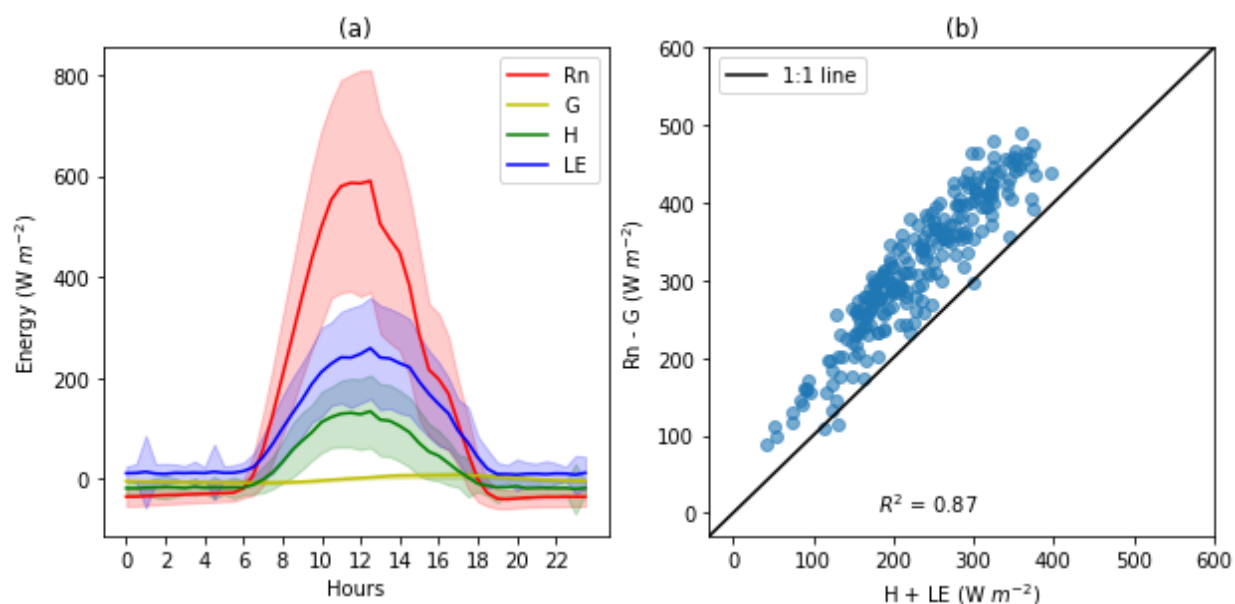
The simulated data was represented using annual, monthly, and daily timescales. The graphical representation included shaded areas representing the range of possible values (standard deviations) due to climate stochasticity and the ecosystem response to it. In addition, boxplots were also used to show the simulated value distributions.

## 3 Results and discussion

### 3.1 Observed flux data for a tropical woodland

The flux tower has been operated since October 2018, however in this study, we are considering the first phase of data collection, which corresponds to 235 days, between October 2018 and June 2019. The energy balance component daily dynamics for the study site is

presented on an hourly-basis (Figure 3a) agreeing with other energy partitioning observations from similar land covers in a tropical environment (Cabral et al., 2015; da Rocha et al., 2009; Giambelluca et al., 2009). The energy balance evaluated here (Figure 3b) showed a higher  $R_n - G$  component than  $H + LE$  as we did not consider the storage flux and the biomass energy storage. In addition, the instruments used to measure  $H + LE$  are different from the ones used for  $R_n - G$ .

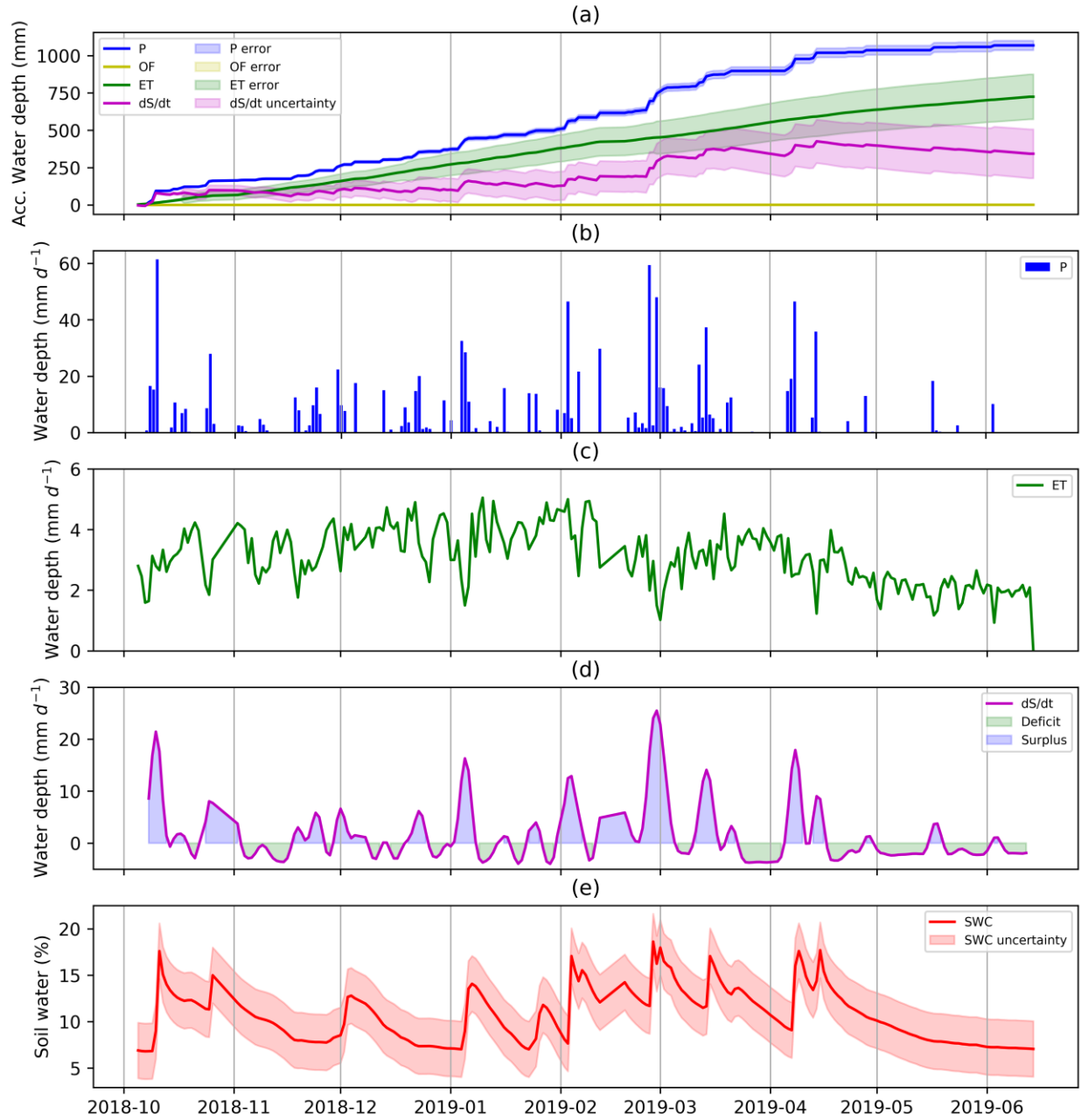


**Figure 3.** (a) energy partitioning throughout a 24-hour period, including net radiation ( $R_n$ ), soil heat flux ( $G$ ), sensible ( $H$ ) and latent ( $LE$ ) heats, the shaded areas that represent the standard deviations of the average calculated every 30 min using the whole dataset; (b) scatter plot between the sum of turbulent fluxes (daily averages) of sensible ( $H$ ) and latent ( $LE$ ) heat versus the available energy ( $R_n - G$ ) (gap filled data included).

The average daily values for the water balance components presented the highest standard deviations for  $P$  measurements (a discrete variable), which reflected on  $dS/dt$  and  $Q$  standard deviations (Table 2). However,  $ET$  presented a lower standard deviation due to its continuity over time. The time series for the main water balance components are presented in Figure 4. The dynamics is similar to what was observed using compatible micrometeorological techniques (Cabral et al., 2015; Valle Júnior et al., 2020). The eddy covariance  $ET$  measurement uncertainties were computed (Table 2 and Figure 4a) and their values were close to the typical error range for this method (Allen, Pereira, Howell, & Jensen, 2011). A low runoff coefficient

led to a small overland flow computed for the monitoring period (Figure 4a) as observed in other hillslope scale studies (Anache et al., 2019; Anache et al., 2017; Guo et al., 2015; Nacinovic et al., 2014). The ET measurement flaws are mainly responsible for the water balance uncertainty, followed by rainfall and overland flow. Previous studies in the same site computed uncertainties higher than 100% for the water balance due to poor ET estimates (relative error of 53%) (Anache et al., 2019). Due to reduced errors from eddy covariance ET measurements (Table 2), we found more reliable results for water partitioning in the wooded Cerrado. This first phase dataset is available at a daily time resolution (Anache et al., 2021).

Throughout the monitoring period, the average observed ET was  $3.12 \pm 0.93 \text{ mm d}^{-1}$  (Table 2), ranging between  $0.93 \text{ mm d}^{-1}$  and  $5.05 \text{ mm d}^{-1}$  (Figure 4c). Other flux measurements performed in similar land cover conditions found values ranging between  $1.0 \text{ mm d}^{-1}$  and  $7.1 \text{ mm d}^{-1}$  (Cabral et al., 2015); averaging  $2.25 \pm 0.45 \text{ mm d}^{-1}$  (Giambelluca et al., 2009); and averaging  $1.36 \text{ mm d}^{-1}$  on dry periods and  $3.08 \text{ mm d}^{-1}$  on wet periods (da Rocha et al., 2009). The evapotranspiration dynamics reflected on the water balance residuals (Figure 4d) similarly to previous studies in the same area (Anache et al., 2019; Oliveira et al., 2015), and it is confirmed by the soil water content (Figure 4e), which presents a similar variation pattern. Additionally, soil water content maintains evapotranspiration rates continuous throughout the dry period (starting in April), agreeing with what was observed by previous field observations in similar ecosystems (Cabral et al., 2015; Giambelluca et al., 2009).



**Figure 4.** Accumulated (a) water balance variables and non-accumulated outcomes: (b) rainfall (c) evapotranspiration, (d) water balance residual, and (e) soil water content for a tropical woodland.

**Table 2.** Summary of observed average daily fluxes in the wooded Cerrado.

Variable	Average value	Standard deviation	Error	
Unit	mm d <sup>-1</sup>	mm d <sup>-1</sup>	mm d <sup>-1</sup>	%
P	4.51	9.78	0.14	3
ET	3.12	0.93	0.64	21
Q	0.004	0.001	0.0004	10
dS/dt	1.46	10.10	0.66	45

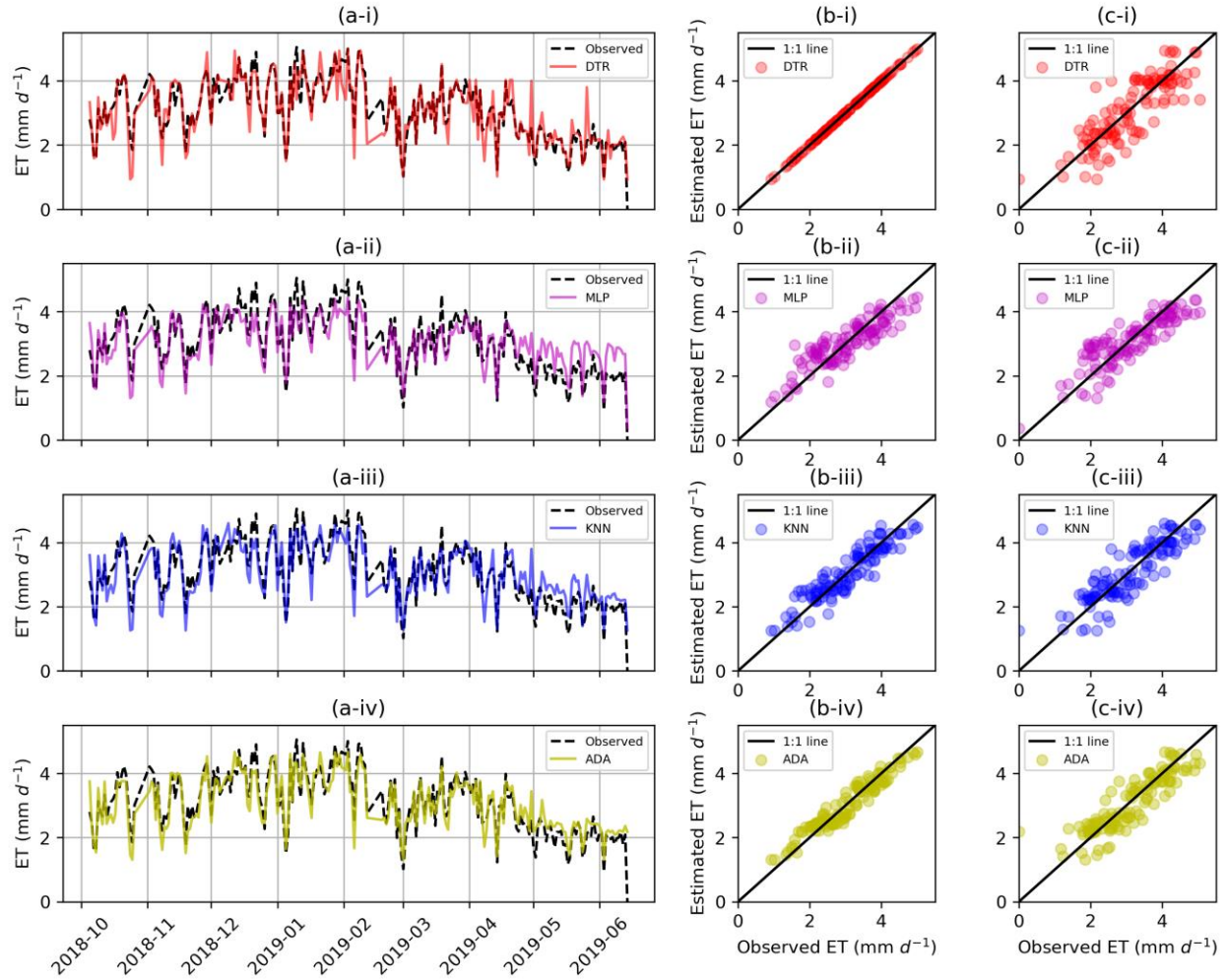
## 3.2 Modeling phase

### 3.2.1 Calibration and validation of models

We ran four machine learning algorithms with training and testing subsets for a 235-day dataset obtained throughout 2018 and 2019. The datasets were composed of a target variable (observed ET, derived from latent heat flux) and explanatory variables, which were used to serve as input to the model: solar radiation (Rg), wind speed (WS), temperature (T), relative humidity (RH), and rainfall (P). These input variables were chosen according to the machine learning modeling strategies discussed by Granata (2019), and we also ranked the input variables according to their importance on estimating ET. Although we were expecting to find radiation as the most relevant variable to model ET (Maček, Bezak, & Šraj, 2018) using machine learning techniques based on linear approaches, we found that T is the most important one, followed by WS, RH, Rg, and P, respectively. We compared the observed data for training and testing subsets (Figures 5b and 5c, respectively) for each machine learning technique. The dispersion is increased on the testing (validation) and all data subsets, as expected, and observed by similar modeling exercises (Dou & Yang, 2018; Goyal et al., 2014; Yamaç & Todorovic, 2020). This behavior is confirmed by the model performance evaluation metrics (Table 3).

The observed evapotranspiration dynamics and range (Figure 5a) were reproduced by all four tested machine learning algorithms throughout the observed period, which included both wet (between October and March) and dry (between April and June) seasons. In general, all models (except ADA) underestimated ET throughout the calibration phase. However, all of them overestimated ET throughout the validation. DTR presented an overfitted calibration (Figure 5b-i), which could not be well reproduced throughout the validation phase (Figure 5c-i), revealing the technique inadequacy to reproduce continuous floating data. Although neural networks are a feasible alternative to model ET (Liu et al., 2012; Yamaç & Todorovic, 2020), KNN and ADA

presented the best performances throughout all the model evaluation phases. The KNN model has been chosen to reproduce ET data throughout the water balance simulation phase.



**Figure 5.** Comparison between observed and estimated ET values: (a) along monitoring time; and using scatter plots for (b) calibration and (c) validation datasets. Machine learning techniques used to estimate ET: (i) decision trees, (ii) neural networks, (iii) k-nearest neighbors, and (iv) AdaBoost predictors.



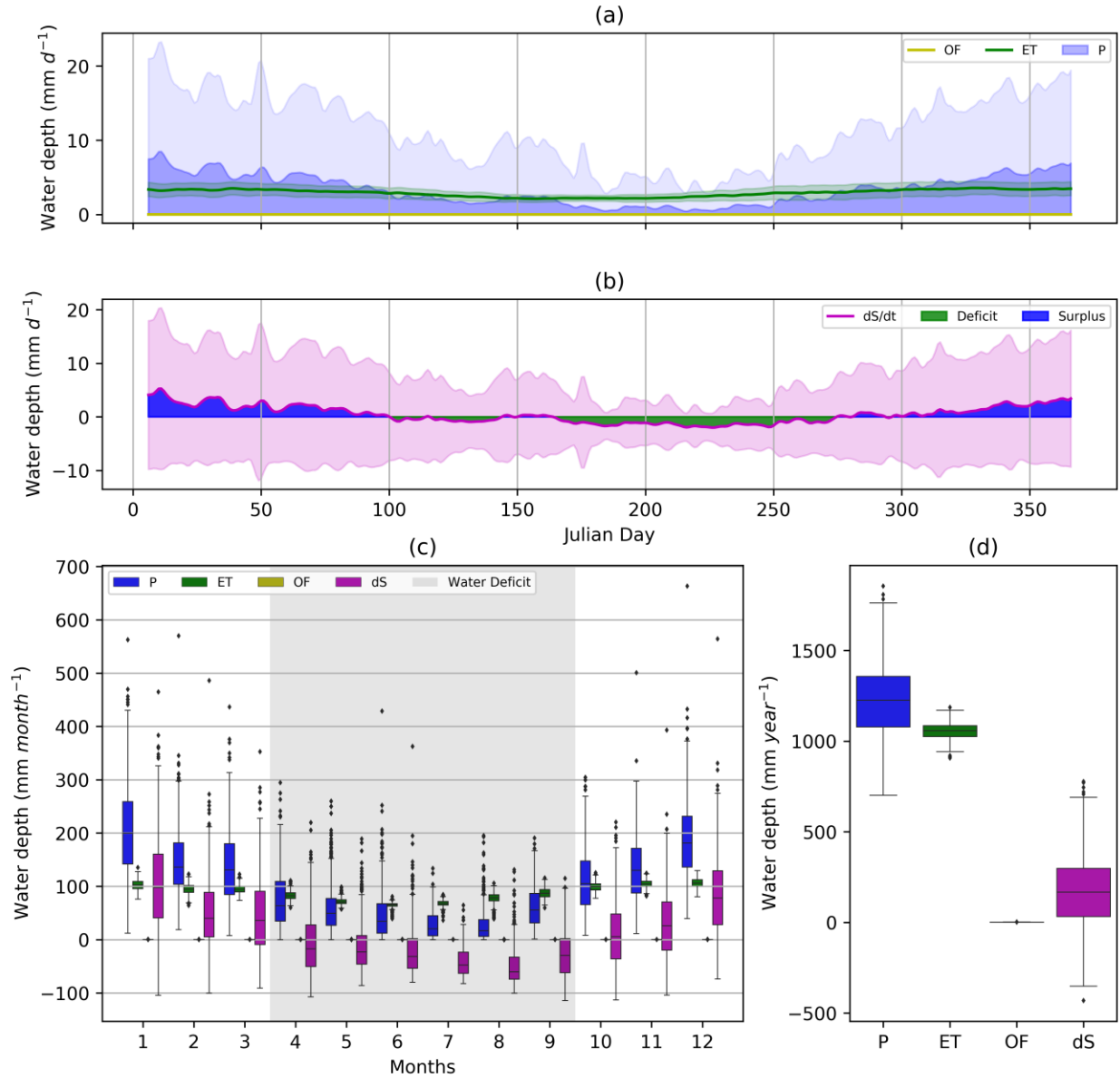
**Table 3.** Performance evaluation of machine learning techniques to predict evapotranspiration for a tropical woodland.

Model	Metrics	Calibration	Validation	All data
DTR	R <sup>2</sup>	1.00	0.65	0.81
	MAE	0.00	0.44	0.22
	KGE	1.00	0.83	0.91
	PBIAS	0.00	-1.51	-0.77
MLP	R <sup>2</sup>	0.77	0.74	0.75
	MAE	0.36	0.40	0.38
	KGE	0.66	0.70	0.69
	PBIAS	0.09	-0.98	-0.45
KNN	R <sup>2</sup>	0.85	0.77	0.81
	MAE	0.29	0.38	0.34
	KGE	0.86	0.84	0.85
	PBIAS	0.37	-0.78	-0.22
ADA	R <sup>2</sup>	0.92	0.77	0.84
	MAE	0.23	0.36	0.30
	KGE	0.85	0.83	0.84
	PBIAS	-0.27	-1.36	-0.82

DTR: decision trees; MLP: neural network multilayer perceptron; KNN: k-nearest neighbors; and ADA: ensemble based AdaBoost. R<sup>2</sup>: coefficient of determination; MSE: mean squared error; PBIAS: percent bias; KGE: Kling-Gupta efficiency. All metrics are dimensionless except MAE (mm d<sup>-1</sup>).

### 3.2.2 Long-term water balance simulation

The long-term simulated water balance presented similar periods of water deficit and surplus at a daily basis (Figure 6a and 6b). Thus, by observing the outcomes, we may identify more periods of water surplus. The simulated ET throughout replicated years representing the local climate stochasticity ranged from 1.25 mm d<sup>-1</sup> to 4.69 mm d<sup>-1</sup>, and the average was 2.91 mm d<sup>-1</sup>.



**Figure 6.** Long-term averages for (a) daily rainfall (P), evapotranspiration (ET), overland flow (OF); (b) water balance residual (dS/dt); shaded areas with the variable color represent its standard deviations; boxplots containing (c) monthly, and (d) annual long-term water balance components distributions: rainfall (P), evapotranspiration (ET), overland flow (OF), and water balance residual (dS); all figures (a, b, c and d) represent the same time series calculated using a stochastic weather generator (KnnCADv4) as input for calibrated OF and ET predictors.

The monthly total average for the water balance components ranged differently between the dry and wet periods (Figure 6c), except for the overland flow, which is not significant (close to zero) compared with the other variables. The driest periods occur between April and

September; and the wet season happens between October and March. Additionally, there is a dependency of the wooded Cerrado water balance in the long term on the rainfall stochasticity. This dependency reflects on the average simulated annual  $dS/dt$  ( $172 \pm 211 \text{ mm yr}^{-1}$ ) and  $P$  ( $1227 \pm 208 \text{ mm yr}^{-1}$ ), which have similar standard deviations (around  $200 \text{ mm yr}^{-1}$ ) (Figure 6d). However, the ET ( $1054 \pm 46 \text{ mm yr}^{-1}$ ) presented high annual rates but a small variability throughout the simulated year, confirming that it depends more on soil water availability than on the rainfall events. This corroborated with the expected moisture cycling dynamics typical from the tropics (Wohl et al., 2012). In addition, it is typical from tropical woodland sites that the ET continues throughout dry periods without rainfall due to the steady transpiration process supported by soil water content (Wang-Erlandsson, van der Ent, Gordon, & Savenije, 2014). Such a steady condition of ET in tropical woodlands is important to maintain regional water balance and moisture cycling between tropics and subtropics in Brazil (aerial rivers) (Nobre, Arraut, Barbosa, Obregon, & Marengo, 2012; Spangler, Lynch, & Spera, 2017; Spera, Galford, Coe, Macedo, & Mustard, 2016), and is supported by an increased soil water retention capacity combined with a deeper root system provided by woody and undisturbed vegetation cover (Dias, Macedo, Costa, Coe, & Neill, 2015; R. S. Oliveira et al., 2005; Tseng, Alves, & Crestana, 2018).

### 3.3 Water partitioning behavior and variability of a tropical woodland

The long-term water balance was averaged in multiple time resolutions (annual, monthly, and daily) (Table 3). This multiple time-scale visualization approach reveals that the climate variability may cause a sort of surplus and scarcity situations. The resilience of the wooded Cerrado to maintain ET throughout the dry periods may cause negative  $dS/dt$  values at multiple timescales. However, the probability of having a dry year ( $dS/dt < 0 \text{ mm yr}^{-1}$ ) is approximately 20%, corresponding to the return period of 5 years, agreeing with previous predictions for drought occurrences in Brazil (Awange, Mpelasoka, & Goncalves, 2016). The probability of a negative  $dS/dt$  increases when considering monthly ( $\sim 50\%$ ) and daily ( $> 90\%$ ) timescales. This information reveals the importance of managing the water resources better, as the study site is located above the Guarani Aquifer System (GAS) outcrop zone, where the infiltrated water that percolates the unsaturated zone may potentially recharge the aquifer (Anache et al., 2019; Lucas & Wendland, 2015; P. T. S. Oliveira et al., 2017).

**Table 4.** Summary of the simulated average annual, monthly, and daily fluxes in the wooded Cerrado.

Timescale	Unit	Variable	Average value	Standard deviation
Annual	mm yr <sup>-1</sup>	P	1227.16	207.63
		ET	1054.01	46.30
		Q	1.23	0.21
		dS/dt	171.93	211.04
Monthly	mm month <sup>-1</sup>	P	102.26	82.46
		ET	87.83	16.30
		Q	0.10	0.08
		dS/dt	14.33	76.71
Daily	mm d <sup>-1</sup>	P	3.39	8.96
		ET	2.91	0.91
		Q	0.003	0.008
		dS/dt	0.47	9.11

#### 4 Conclusions

This paper presented new water flux measurements in an undisturbed environment of the Brazilian Cerrado: a tropical woodland, also known as wooded Cerrado. We enhanced experimental data visualization and exploration using computational and modeling techniques. The long-term hydrological simulation to compute the water balance components was necessary to better explore and expand the possibilities for a window of flux measurements in the field. These experimental data alone provide us with particularly important information about the water partitioning on tropical woodlands, nevertheless, they do not reveal the whole story. Consequently, using machine learning and stochastic simulations were necessary to complete the information to understand the system better, by analyzing the flux variabilities on multiple timescales and including climate stochasticity.

The daily water balance of the wooded Cerrado in both observed and long-term prediction datasets presented high dependence on daily rainfall occurrence, as it is a discrete variable and varies significantly throughout the days. Conversely, evapotranspiration has a continuous characteristic, leading to a smaller variation throughout time, smoothing the ecosystem responses written in the water balance residuals and soil water content variations. In addition, here we confirmed that wooded Cerrado vegetation supports the evapotranspiration by

transferring water from soil moisture to the atmosphere, and rainfall variation has a low impact on this variable.

Considering the long-term simulated water balance series, a resilient behavior to droughts could be identified in the wooded Cerrado, as it tends to have 6 months of water surplus and 6 months of water deficit throughout the simulated years, and the latter was slightly lower. This shows the importance of water storage in the subsurface throughout the dry season, which helps to maintain the moisture cycling over the whole year for this ecosystem and surrounding areas. The wooded Cerrado water balance in the long term also relies on the rainfall stochasticity. This dependency can be observed on the average simulated annual  $dS/dt$  (water balance residual) ( $172 \pm 211 \text{ mm yr}^{-1}$ ) and  $P$  ( $1227 \pm 208 \text{ mm yr}^{-1}$ ), which have similar standard deviations (around  $200 \text{ mm yr}^{-1}$ ); different from the  $ET$  ( $1054 \pm 46 \text{ mm yr}^{-1}$ ), which presented high annual rates but a small variability throughout the simulated years, not significantly reflecting the water balance residual. In addition, the probability of occurrence of a dry year ( $dS/dt < 0$ ) is close to 20%, revealing a return period of approximately 5 years to droughts.

## Acknowledgements and data

This study was supported by grants from the Ministry of Science, Technology, and Innovation of Brazil – MCTI – and the National Council for Scientific and Technological Development – CNPq (grant number 150057/2018-0); the São Paulo Research Foundation – FAPESP (grant number 2015/03806-1); and the Coordination of Improvement of Higher Education Personnel – CAPES (finance code 001).

The authors acknowledge the graduate program in Hydraulics and Sanitary Engineering – PPGSHS (USP-EESC) – for the scientific support and the Arruda Botelho Institute – IAB – for allowing the development of this study on its private land. The authors would also like to thank the editor and anonymous referees for their useful comments, which substantially improved the paper.

Datasets for this research are available in these in-text data citation reference: Anache et al. (2021) (daily hydrometeorological observations and machine learning training and testing dataset); and Marin et al. (2021) (weather generator inputs).

## References

- Allard, D., Ailliot, P., Monbet, V., & Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, 156(1), 101-113.
- Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, 98(6), 899-920. doi:10.1016/j.agwat.2010.12.015
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. Rome: FAO - Food and Agriculture Organization of the United Nations.
- Anache, J. A. A., Flanagan, D. C., Srivastava, A., & Wendland, E. C. (2018). Land use and climate change impacts on runoff and soil erosion at the hillslope scale in the Brazilian Cerrado. *Sci Total Environ*, 622-623, 140-151. doi:10.1016/j.scitotenv.2017.11.257
- Anache, J. A. A., Rosalem, L. M. P., Ballarin, A. S., Kobayashi, A. N. A., Santos, A. C., & Wendland, E. (2021). Supplement for "Tropical woodland moisture cycling responses to climate stochasticity using eddy covariance and machine learning". doi:10.5281/zenodo.4604102
- Anache, J. A. A., Wendland, E., Rosalem, L. M. P., Youlton, C., & Oliveira, P. T. S. (2019). Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. *Hydrology and Earth System Sciences*, 23(3), 1263-1279. doi:10.5194/hess-23-1263-2019
- Anache, J. A. A., Wendland, E. C., Oliveira, P. T. S., Flanagan, D. C., & Nearing, M. A. (2017). Runoff and soil erosion plot-scale studies under natural rainfall: A meta-analysis of the Brazilian experience. *Catena*, 152, 29-39. doi:10.1016/j.catena.2017.01.003
- Apipattanas, S., Podestá, G., Rajagopalan, B., & Katz, R. W. (2007). A semiparametric multivariate and multisite weather generator. *Water Resources Research*, 43(11). doi:10.1029/2006wr005714
- Awange, J. L., Mpelasoka, F., & Goncalves, R. M. (2016). When every drop counts: Analysis of Droughts in Brazil for the 1901-2013 period. *Sci Total Environ*, 566-567, 1472-1488. doi:10.1016/j.scitotenv.2016.06.031
- Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Glob Chang Biol*, 26(1), 242-260. doi:10.1111/gcb.14807
- Bastiaanssen, W. G. M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J. F., Roerink, G. J., & van der Wal, T. (1998). A remote sensing surface energy balance algorithm for land (SEBAL) 2. Validation. *Journal of Hydrology*, 212-213, 213-229.
- Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., . . . Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, 64(10), 1141-1158. doi:10.1080/02626667.2019.1620507
- Burt, T. P., & McDonnell, J. J. (2015). Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resour Res*, 51(8), 5919-5928. doi:10.1002/2014WR016839
- Cabral, O. M. R., da Rocha, H. R., Gash, J. H., Freitas, H. C., & Ligo, M. A. V. (2015). Water and energy fluxes from a woodland savanna (cerrado) in southeast Brazil. *J Hydrol: Regional Studies*, 4, 22-40. doi:10.1016/j.ejrh.2015.04.010
- Cabrera, M. C. M., Anache, J. A. A., Youlton, C., & Wendland, E. (2016). Performance of evaporation estimation methods compared with standard 20 m<sup>2</sup> tank. *Rev Bras Eng Agr Amb*, 20(10), 874-879. doi:10.1590/1807-1929/agriambi.v20n10p874-879
- da Rocha, H. R., Manzi, A. O., Cabral, O. M., Miller, S. D., Goulden, M. L., Saleska, S. R., . . . Maia, J. F. (2009). Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil. *J Geophys Res*, 114(G1). doi:10.1029/2007jg000640
- Dias, L. C. P., Macedo, M. N., Costa, M. H., Coe, M. T., & Neill, C. (2015). Effects of land cover change on evapotranspiration and streamflow of small catchments in the Upper Xingu River Basin, Central Brazil. *J Hydrol: Regional Studies*, 4, 108-122. doi:10.1016/j.ejrh.2015.05.010
- Dijk, A. v. (2002). Extension to 3D of "The Effect of Line Averaging on Scalar Flux Measurements with a Sonic Anemometer near the Surface" by Kristensen and Fitzjarrald. *Journal of Atmospheric and Oceanic Technology*, 19(1), 80-82. doi:10.1175/1520-0426(2002)019<0080:etoteo>2.0.co;2
- Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture*, 148, 95-106. doi:10.1016/j.compag.2018.03.010

- Foken, T., Leuning, R., Oncley, S. R., Mauder, M., & Aubinet, M. (2012). Corrections and Data Quality Control. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Dordrecht Heidelberg London New York: Springer.
- Giambelluca, T. W., Scholz, F. G., Bucci, S. J., Meinzer, F. C., Goldstein, G., Hoffmann, W. A., . . . Buchert, M. P. (2009). Evapotranspiration and energy balance of Brazilian savannas with contrasting tree density. *Agr Forest Meteorol*, 149(8), 1365-1376. doi:10.1016/j.agrformet.2009.03.006
- Goyal, M. K., Bharti, B., Quilty, J., Adamowski, J., & Pandey, A. (2014). Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Systems with Applications*, 41(11), 5267-5276. doi:10.1016/j.eswa.2014.02.047
- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, 217, 303-315. doi:10.1016/j.agwat.2019.03.015
- Guo, Q., Hao, Y., & Liu, B. (2015). Rates of soil erosion in China: A study based on runoff plot data. *Catena*, 124, 68-76. doi:10.1016/j.catena.2014.08.013
- Herrera, M., Natarajan, S., Coley, D. A., Kershaw, T., Ramallo-González, A. P., Eames, M., . . . Wood, M. (2017). A review of current and future weather data for building simulation. *Building Services Engineering Research and Technology*, 38(5), 602-627. doi:10.1177/0143624417705937
- Horst, T. W., & Lenschow, D. H. (2009). Attenuation of Scalar Fluxes Measured with Spatially-displaced Sensors. *Boundary-Layer Meteorology*, 130(2), 275-300. doi:10.1007/s10546-008-9348-0
- Kaimal, J. C., Clifford, S. F., & Lataitis, R. J. (1989). Effect of finite sampling on atmospheric spectra. *Boundary-Layer Meteorology*, 47(1), 337-347. doi:10.1007/bf00122338
- King, L. M., McLeod, A. I., & Simonovic, S. P. (2015). Improved Weather Generator Algorithm for Multisite Simulation of Precipitation and Temperature. *JAWRA Journal of the American Water Resources Association*, 51(5), 1305-1320. doi:10.1111/1752-1688.12307
- Kwon, H.-H., Sivakumar, B., Moon, Y.-I., & Kim, B.-S. (2010). Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. *Stochastic Environmental Research and Risk Assessment*, 25(4), 567-581. doi:10.1007/s00477-010-0422-z
- Levis, C., Flores, B. M., Mazzochini, G. G., Manhaes, A. P., Campos-Silva, J. V., Borges de Amorim, P., . . . Clement, C. R. (2020). Help restore Brazil's governance of globally important ecosystem services. *Nat Ecol Evol*, 4(2), 172-173. doi:10.1038/s41559-019-1093-x
- Liu, Z., Peng, C., Xiang, W., Deng, X., Tian, D., Zhao, M., & Yu, G. (2012). Simulations of runoff and evapotranspiration in Chinese fir plantation ecosystems using artificial neural networks. *Ecological Modelling*, 226, 71-76. doi:10.1016/j.ecolmodel.2011.11.023
- Lucas, M., & Wendland, E. (2015). Recharge estimates for various land uses in the Guarani Aquifer System outcrop area. *Hydrolog Sci J*, 61(7), 1253-1262. doi:10.1080/02626667.2015.1031760
- Maček, U., Bezak, N., & Šraj, M. (2018). Reference evapotranspiration changes in Slovenia, Europe. *Agricultural and Forest Meteorology*, 260-261, 183-192. doi:10.1016/j.agrformet.2018.06.014
- Mandal, S., Arunkumar, R., Breach, P. A., & Simonovic, S. P. (2019). Reservoir Operations under Changing Climate Conditions: Hydropower-Production Perspective. *Journal of Water Resources Planning and Management*, 145(5), 04019016. doi:10.1061/(asce)wr.1943-5452.0001061
- Marin, F. R., Pilau, F. G., & Sentelhas, P. C. (2021). Posto Meteorológico "Professor Jesus Marden dos Santos" ESALQ - USP. <http://www.leb.esalq.usp.br/posto/index.html>
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., . . . Verhoest, N. E. C. (2017). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5), 1903-1925. doi:10.5194/gmd-10-1903-2017
- McCabe, M. F., & Wood, E. F. (2006). Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sensing of Environment*, 105(4), 271-285. doi:10.1016/j.rse.2006.07.006
- Melo, D. C. D., Anache, J. A. A., Almeida, C. d. N., Coutinho, J. V., Ramos Filho, G. M., Rosalem, L. M. P., . . . Wendland, E. (2020). The big picture of field hydrology studies in Brazil. *Hydrological Sciences Journal*, null-null. doi:10.1080/02626667.2020.1747618
- Moncrieff, J. B., Massheder, J. M., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., . . . Verhoef, A. (1997). A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *Journal of Hydrology*, 188, 589-611. doi:10.1016/S0022-1694(96)03194-0
- Montgomery, R. B. (1947). Viscosity and thermal conductivity of air and diffusivity of water vapor in air. *Journal of Meteorology*, 4(6), 193-196. doi:10.1175/1520-0469(1947)004<0193:vatcoa>2.0.co;2
- Moore, C. J. (1986). Frequency response corrections for eddy correlation systems. *Boundary-Layer Meteorology*, 37(1), 17-35. doi:10.1007/bf00122754

- Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8), 1781-1800. doi:10.1016/j.rse.2011.02.019
- Nacinovic, M. G. G., Mahler, C. F., & Avelar, A. d. S. (2014). Soil erosion as a function of different agricultural land use in Rio de Janeiro. *Soil Till Res*, 144, 164-173. doi:10.1016/j.still.2014.07.002
- Nicks, A. D., Lane, L. J., & Gander, G. A. (1995). Chapter 2. Weather Generator. In D. Flanagan & M. A. Nearing (Eds.), *Hillslope profile and watershed model documentation. NSERL Report No. 10*. West Lafayette, Indiana: USDA-ARS National Soil Erosion Research Laboratory.
- Nobre, C., Arraut, J. M., Barbosa, H. M. J., Obregon, G., & Marengo, J. (2012). Aerial Rivers and Lakes: Looking at Large-Scale Moisture Transport and Its Relation to Amazonia and to Subtropical Rainfall in South America. *Journal of Climate*, 25(2), 543-556. doi:10.1175/2011jcli4189.1
- Nobrega, R. L. B., Guzha, A. C., Torres, G. N., Kovacs, K., Lamparter, G., Amorim, R. S. S., . . . Gerold, G. (2017). Effects of conversion of native cerrado vegetation to pasture on soil hydro-physical properties, evapotranspiration and streamflow on the Amazonian agricultural frontier. *PLoS ONE*, 12(6), e0179414. doi:10.1371/journal.pone.0179414
- Oliveira, P. T. S., Leite, M. B., Mattos, T., Nearing, M. A., Scott, R. L., Xavier, R. O., . . . Wendland, E. (2017). Groundwater recharge decrease with increased vegetation density in the Brazilian cerrado. *Ecohydrology*, 10(1), e1759. doi:10.1002/eco.1759
- Oliveira, P. T. S., Nearing, M. A., Hawkins, R. H., Stone, J. J., Rodrigues, D. B. B., Panachuki, E., & Wendland, E. (2016). Curve number estimation from Brazilian Cerrado rainfall and runoff data. *J Soil Water Conserv*, 71(5), 420-429. doi:10.2489/jswc.71.5.420
- Oliveira, P. T. S., Nearing, M. A., Moran, M. S., Goodrich, D. C., Wendland, E., & Gupta, H. V. (2014). Trends in water balance components across the Brazilian Cerrado. *Water Resour Res*, 50(9), 7100-7114. doi:10.1002/2013WR015202
- Oliveira, P. T. S., Wendland, E., Nearing, M. A., Scott, R. L., Rosolem, R., & da Rocha, H. R. (2015). The water balance components of undisturbed tropical woodlands in the Brazilian cerrado. *Hydrol Earth Syst Sc*, 19(6), 2899-2910. doi:10.5194/hess-19-2899-2015
- Oliveira, R. S., Bezerra, L., Davidson, E. A., Pinto, F., Klink, C. A., Nepstad, D. C., & Moreira, A. (2005). Deep root function in soil water dynamics in cerrado savannas of central Brazil. *Funct Ecol*, 19(4), 574-581. doi:10.1111/j.1365-2435.2005.01003.x
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., . . . Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11), 933-938. doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 11(5), 1633-1644. doi:10.5194/hess-11-1633-2007
- Priestley, C. H. B., & Taylor, R. J. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Mon Weather Rev*, 100(2), 81-92. doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2
- Richardson, C., & Wright, D. (1984). *WGEN: a model for generating daily weather variables*. Washington, DC: US Department of Agriculture, Agricultural Research Service.
- Ruhoff, A. L., Paz, A. R., Aragao, L. E. O. C., Mu, Q., Malhi, Y., Collischonn, W., . . . Running, S. W. (2013). Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin. *Hydrological Sciences Journal*, 58(8), 1658-1676. doi:10.1080/02626667.2013.837578
- Ruhoff, A. L., Paz, A. R., Collischonn, W., Aragao, L. E. O. C., Rocha, H. R., & Malhi, Y. S. (2012). A MODIS-Based Energy Balance to Estimate Evapotranspiration for Clear-Sky Days in Brazilian Tropical Savannas. *Remote Sensing*, 4(3), 703-725. doi:10.3390/rs4030703
- Schotanus, P., Nieuwstadt, F. T. M., & De Bruin, H. A. R. (1983). Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes. *Boundary-Layer Meteorology*, 26(1), 81-93. doi:10.1007/bf00164332
- Semenov, M. A., Brooks, R. J., Barrow, E. M., & Richardson, C. W. (1998). Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, 10(2), 95-107.
- Semenov, M. A., & Stratonovitch, P. (2010). Use of multi-model ensembles from global climate models for assessment of climate change impacts. *Climate Research*, 41, 1-14. doi:10.3354/cr00836



- Shapland, T. M., Snyder, R. L., Paw U, K. T., & McElrone, A. J. (2014). Thermocouple frequency response compensation leads to convergence of the surface renewal alpha calibration. *Agricultural and Forest Meteorology*, 189-190, 36-47. doi:10.1016/j.agrformet.2014.01.008
- Spangler, K. R., Lynch, A. H., & Spera, S. A. (2017). Precipitation Drivers of Cropping Frequency in the Brazilian Cerrado: Evidence and Implications for Decision-Making. *Weather, Climate, and Society*, 9(2), 201-213. doi:10.1175/wcas-d-16-0024.1
- Spera, S. A., Galford, G. L., Coe, M. T., Macedo, M. N., & Mustard, J. F. (2016). Land-use change affects water recycling in Brazil's last agricultural frontier. *Glob Chang Biol*, 22(10), 3405-3413. doi:10.1111/gcb.13298
- Talsma, C., Good, S., Miralles, D., Fisher, J., Martens, B., Jimenez, C., & Purdy, A. (2018). Sensitivity of Evapotranspiration Components in Remote Sensing-Based Models. *Remote Sensing*, 10(10), 1601. doi:10.3390/rs10101601
- Tanner, C. B., & Thurtell, G. W. (1969). Anemoclinometer measurements of Reynolds stress and heat transport in the atmospheric surface layer science lab. *US Army Electronics Command - Atmospheric Sciences Laboratory, TR ECOM 66-G22-F*, R1-R10.
- Tseng, C. L., Alves, M. C., & Crestana, S. (2018). Quantifying physical and structural soil properties using X-ray microtomography. *Geoderma*, 318, 78-87. doi:10.1016/j.geoderma.2017.11.042
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, 11(5), 910. doi:10.3390/w11050910
- Valle Júnior, L. C. G., Ventura, T. M., Gomes, R. S. R., de S. Nogueira, J., de A. Lobo, F., Vourlitis, G. L., & Rodrigues, T. R. (2020). Comparative assessment of modelled and empirical reference evapotranspiration methods for a brazilian savanna. *Agricultural Water Management*, 232, 106040. doi:10.1016/j.agwat.2020.106040
- Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J., & Savenije, H. H. G. (2014). Contrasting roles of interception and transpiration in the hydrological cycle – Part 1: Temporal characteristics over land. *Earth System Dynamics*, 5(2), 441-469. doi:10.5194/esd-5-441-2014
- Wang, K., & Dickinson, R. E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, 50(2). doi:10.1029/2011rg000373
- Webb, E. K., Pearman, G. I., & Leuning, R. (1980). Correction of flux measurements for density effects due to heat and water vapour transfer. *Quarterly Journal of the Royal Meteorological Society*, 106(447), 85-100. doi:10.1002/qj.49710644707
- Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., . . . Ogden, F. (2012). The hydrology of the humid tropics. *Nat Clim Change*, 2(9), 655-662. doi:10.1038/nclimate1556
- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., . . . Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, 15(16), 5015-5030. doi:10.5194/bg-15-5015-2018
- Yamaç, S. S., & Todorovic, M. (2020). Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agricultural Water Management*, 228, 105875. doi:10.1016/j.agwat.2019.105875
- Yazd, H. R. G. H., Salehnia, N., Kolsoumi, S., & Hoogenboom, G. (2019). Prediction of climate variables by comparing the knearest neighbor method and MIROC5 outputs in an arid environment. *Climate Research*, 77(2), 99-114. doi:10.3354/cr01545