Calibration and error investigation of large tipping bucket flow meters

Abstract

Inherent errors in tipping bucket flow meters may limit monitoring data reliability. In this work, we perform the static and dynamic calibration of four large tipping buckets, apply different regression curves and investigate the possible measurement error sources. The volumetric capacity (static calibration) of each piece of equipment was determined. They were tested (dynamic calibration) under ten flow intensities, ranging from low to high rainfall intensities (return period larger than 100 years). For each flow rate, the measurement was recorded during six time intervals (1, 2, 5, 10, 20 and 30 minutes) and four regression equations - linear, potential, T vs. 1/Q and quadratic - were tested. According to the static calibration, the equipment has a volumetric capacity of 11.63 mL (TB1), 64.16 mL (TB2), 139.86 mL (TB3) and 660.95 mL (TB4). When tested under different flow rates (dynamic calibration), underestimations were identified according to the size of the cavity: TB1 (3.31%), TB2 (5.75%), TB3 (9.33%) and TB4 (13.57%). Among the alternative curves, linear regression showed the best correlation (above 99%) with the monitored data. Using this method, the measurement errors were reduced to -1.35% (TB1), 0.04% (TB2), 3.18% (TB3) and 3.73% (TB4). We investigated how the different variables (tipping speed, cavity volumetric capacity and time interval of data collection) influenced the error. Errors follow a parabolic function of tipping velocity and a linear function of cavity volumetric capacity. The time interval of data collection interfered in the data sampled, however no statistical correlation was found. Among those variables, cavity size is the most important one. Given its low cost we aimed to minimize the inherent error in large tipping buckets flow meters and encourage its application, increasing in-situ collection of hydrological data.

Keywords: systemic errors; runoff; in-situ monitoring; hillslope hydrology; land use and land cover; device.

1 Introduction

Rapid land use and land cover (LULC) changes (Chanapathi & Thatikonda, 2020; Mello *et al.*, 2020), climate change (Rocha *et al.*, 2020; Yang *et al.*, 2020) and population growth (Kifle Arsiso *et al.*, 2017) result in higher demand of water, food and energy (Mahlknecht *et al.*, 2020). Through in-situ monitoring, long-term data are created to support the development of new technologies and solutions to maintain the hydrological cycle (Nóbrega *et al.*, 2017; Anache *et al.*, 2019).

Considering the specific requirements for hydrological monitoring in a study area, there are many alternative instruments available, while each one has its own advantage and limitations (Sun *et al.*, 2014). A tipping bucket (TB) flow meter is a robust, simple and high mobility monitoring piece of equipment, which is easy to install and maintain (Sun *et al.*, 2014; Shimizu *et al.*, 2018). A TB consists of two cavities divided by a vertical plate at stable positions (one remains up while the other is down) and it uses a very simple operational mechanism: the collected water flow falls into a cavity and once it reaches its volumetric capacity, the gravitational mass center is switched towards the full cavity, rising the empty cavity and replacing the previous one, while releasing the stored water. This process repeats during the entire flow event. In addition to monitoring accurately, using this kind of equipment when associated with a reed switch and datalogger allows automation and better details of the data collected, such as identifying the beginning, end, and peak of the flow (Corona *et al.*, 2013; Sun *et al.*, 2014; Zabret *et al.*, 2018).

Despite the fact that tipping bucket application dates back to 1928 (Nebol'sin, 1928), it has been used for surface/subsurface flows in small study area measurements, such as runoff (Calder and Kidd, 1978; Chow, 1976; Corona *et al.*, 2013; Elder *et al.*, 2014; Hollis & Ovenden, 1987; Khan & Ong, 1997; Klik *et al.*, 2004; Kim *et al.*, 2005; Johnston, 1942; Nehls *et al.*, 2011; Peyrard *et al.*, 2016; Perales-Momparler *et al.*, 2017; Langhans *et al.*, 2019; Wang *et al.*, 2020; Whipkey, 1965), percolation (Peyrard *et al.*, 2016; Lamb *et al.*, 2019; Wang *et al.*, 2020), throughfall (Takahashi *et al.*, 2010; Zabret *et al.*, 2018) and stemflow (Takahashi *et al.*, 2010; Iida *et al.*, 2012; Shimizu *et al.*, 2018; Zabret *et al.*, 2018).

As well as tipping bucket rain gauges, tipping bucket flow meters used for runoff measurements are also susceptible to measurement errors between the reference and measured flows, thus requiring the application of calibration curves. Calibration can be done in two ways: static (volumetric) and dynamic. Static calibration consists of determining the volume of water necessary for the center of mass to be shifted towards the filling cavity, leading to its tipping. The volume determined in this step corresponds to the equipment's reference volume or its volumetric capacity. On the other hand, dynamic calibration consists of plotting sample points in a graph, which correlates reference and measured flows and then using regression curves to minimize errors. Unlike the static calibration that occurs under extremely low flow rates, usually by drops, to minimize the kinetic energy of the water, in dynamic calibration, the measurements occur under different flow rates (Shedekar *et al.*, 2016). Further details about both methodologies mentioned are given by Humphrey *et al.* (1997).

After lab tests and data collection for dynamic calibration, another phase begins: application of regression curves that best fit the data. There is a large number of applicable equations, ranging from the simplest (linear) to the most complex (polynomial and exponential). Calder and Kidd (1978) identified a non-linearity of errors under increased flow, and thus proposed a new calibration curve by correlating the input flow and time interval between tilts. Based on the same central idea of describing the errors considering its non-linearity, other authors have also proposed calibration curves (Iida *et al.*, 2012; Shimizu *et al.*, 2018; Shiraki & Yamato, 2004; Takahashi *et al.*, 2010).

Despite its recognized applicability, using TBs has systemic errors that need minimization, through calibration, for a more accurate estimation of water flow. Edwards *et al.* (1974) were pioneers in error investigation and development of calibration curves for large TBs. They discovered that water kinetic energy and the volume lost during cavity switching, after reaching volumetric capacity, are some of the main sources of errors in TBs. Since then, many others have dedicated their time to developing calibration techniques (Calder & Kidd, 1978; Iida *et al.*, 2012; Shimizu *et al.*, 2018), while others have focused on applying existing methods and investigating the sources of the errors (Barfield and Hirschi, 1986; Calder & Kidd, 1978; Edwards *et al.*, 1974; Egorov *et al.*, 2015; Hollis & Ovenden, 1987; Iida *et al.*, 2012; Kanzari *et al.*, 2018; Shimizu *et al.*, 2018; Langhans *et al.*, 2019; Sun *et al.*, 2014; Somavilla *et al.*, 2019; Takahashi *et al.*, 2010; Yahaya *et al.*, 2009). Nowadays, Shimizu *et al.* (2018), which is one of the most outstanding papers, provides a general calibration equation for TBs with flat triangular buckets. Although it was successful in eliminating the 2–3% errors in stemflow measurements, it is only applicable for low flow rates (less than 60 mL per minute), inapplicable for most surface flow measurements.

Errors in TBs can be significantly reduced by static and dynamic calibrations (Shedekar *et al.*, 2016), but some observed errors are still not completely minimized (Shimizu *et al.*, 2018). In this context, some questions remain unclear: How can errors be affected by main operational and design variables (tipping velocity, cavity size and time interval)? Among the existing regression curves, which is the most suitable for minimizing errors? Is there a pattern in occurring errors in TBs? Based on these questions, this paper uses different techniques (static and dynamic calibration) and regression curves (linear, potential, T vs. 1/Q and quadratic) aiming to minimize and investigate the source of the occurring errors in four large sizes of tipping buckets flow meters.

2 Methodology

2.1 Tipping bucket description

The tipping bucket flow meters (Figure 1) were designed to measure runoff in the outlet of experimental plots (100 m²) under four different LULC: Wooded Cerrado, also known as Cerrado sensu stricto (TB1), sugarcane (TB2), pasture (TB3) and bare soil (TB4). The plots, which have been operating since 2011 (Youlton *et al.*, 2016a, 2016b; Anache *et al.*, 2018, 2019), are located at the Arruda Botelho Institute, Itirapina, central region of the State of São Paulo – Brazil (latitude 22º10'S, longitude 47º52'O, elevation of 790 m).

INSERT FIGURE 1

2.2 Calibration techniques

To construct an adequate calibration curve, the conditions to be found in the field were evaluated, as the equipment will be applied in determining the runoff in natural and agricultural areas, and is therefore susceptible to the presence of sediments. A high concentration of sediments can influence the water density, as well as accumulate in the cavities of the equipment (Egorov *et al.*, 2015; Langhans *et al.*, 2019). In both cases, they result in malfunction and measurement errors.

Through the construction of a histogram of the concentration of sediments occurring in the study area from 2011 to 2017, it was identified that the highest concentration recorded in the period was 10.2 grams per liter, while most of the events monitored (95%) have a concentration of around 3.0 g/L. Barfield and Hirschi (1986) carried out the calibration process of four scales used to measure surface flow under different concentrations of sediments and concluded that at a concentration below 20 g/L, the presence of sediments can be neglected. Therefore, the adverse effects mentioned above due to the presence of sediment were disregarded, and thus water from the public supply system was used instead of a mixture of water and soil. Figure 2 shows an illustration of the methodological process applied during the static (a) and dynamic (b) calibration processes. A better description of each calibration is given below.

INSERT FIGURE 2

2.2.1 Static calibration

Prior to the calibration step, both cavities must have the same, or as similar as possible, water storage capacity. Thus, preliminary tests were carried out to ensure this consideration by increasing or decreasing the height of the adjustment bar. To determine the volumetric resolution or nominal volume (NV), a graduated pipette and a pipette bulb were used. The water was dripped slowly (interval greater than 2 seconds) so that the kinetic effect did not interfere in the process until one of the cavities tipped and the volume was identified. The procedure was performed ten times in each cavity and then the average between the measurements was applied to determine the equipment's NV.

2.2.2 Dynamic calibration

During the dynamic calibration process, we used a water column made of a PVC pipe with 250 mm of diameter and 1.5 meters in height kept at a constant water level and hydraulic head. In the apparatus, water from the public supply system provides water to the interior of the PVC pipe, keeping the water level constant by overflowing the pipe. A valve at the base of the tube allowed water to escape and enter the equipment's cavities. The reed switch previously installed on the TB, coupled to a datalogger (*Campbell Scientific Inc* CR10 and measuring at 1-minute interval) and a 12V battery, allowed the counting and automatic recording of the number of tips.

In order to test the equipment's behavior under extreme conditions, TBs were tested under runoff rates corresponding to different rainfall return periods. Flow rates were estimated using an Intensity-Duration-Frequency curve - IDF (Equation 1) (Rosalem *et al.*, 2018). The IDF curve was obtained from 40 years of daily precipitation data from the meteorological monitoring station located at the Center for Water Resources and Environmental Studies (CRHEA) at the University of São Paulo, located 5 km far from the application area.

$$I = 1249.\frac{T^{0.15}}{(t+11.39)^{0.81}}$$ Equation **1**

Where: I is the average rainfall intensity (mm.h^{-1}) associated with a return period T (years) and duration t (minutes) adopted.

To construct the dynamic calibration curve of each TB, ten sampling points (runoff flow rates) were calculated based on different return periods and uniformly distributed, with the last sampled flow point resulting from precipitation with a return period greater than 100 years.

The water that flows into the TB (reference flow) was determined by gravimetry, in which the mass of water reserved over a minute was measured on a precision scale or electronic scale, when the maximum measurement limit of the precision scale was reached.

This procedure was carried out in three replicates, at the start and end of each sampling, where the average of these six values was used for the final determination of the reference flow.

In order to investigate the sampling time length interference on measurements, the data collected were grouped into six-time intervals: 1, 2, 5, 10, 20 and 30 minutes. To reduce the possibility of interference from adverse effects, measurements were made in five replicates for each time interval, and the average of the replicates was subsequently calculated.

For the regression curve application, the volume and flow measured by the equipment (called simulated volume and flow here) need to be determined. The simulated volume is the product between the nominal volume and the number of dumps measured in the time interval under analysis, while the simulated flow is the quotient between the simulated volume and the time interval.

2.2.3 Calibration curves

The second part of the dynamic calibration process consists of applying mathematical and statistical techniques searching for equations that best predict simulated and reference flows. To do this, we used four equations: linear; potential; time as function of the inverse of flow rates, and quadratic. Data representation techniques were applied to each of the four TBs under the different time intervals, totaling 96 adjustment curves. A better description of each curve is given below.

A simple linear regression (Equation 2) was the first option used to establish a relationship between the reference (x-axis) and the simulated (y-axis) flows. Other authors (Shimizu *et al.*, 2018) have investigated errors in TBs and mention that the correlation may not be linear. Thus, we investigated the error behaviors under non-linear functions, in which the potential (Equation 2) is one of them. We did not include an intercept value in a linear nor a potential curve, since it would represent a simulated flow associated with a null reference flow. The third regression curve consisted of the graphical representation of the time between dumps (T) as a function of the inverse of the reference flow (Q_{ref}^{-1}), as suggested by Calder and Kidd (1978), called T vs. 1/Q curve here. The time between dumps was calculated by the quotient between the time interval and the number of dumps registered (Equation 2), which was plotted against the inverse of the reference flow, resulting in the mathematical representation of the regression used given by Equation 2. Finally, a quadratic model regression was applied between the dump volume (V_{tip}) as a function of the reference flow (Q_{ref}), as proposed by Costello and Williams (1991). This regression technique assumes that there is a change in the nominal volume according to the reference flow rate. Once the instantaneous tipping volume is calculated (Equation 2), the simulated flow can be estimated by this mathematical representation (Equation 2).

$$\imath Q_{ref}$$ Equation 2

$$Q_{bas} = a Q_{ref}^{b}$$ Equation 3

$$T = \Delta t / n$$ Equation 4

$$T = V_{basc} Q_{ref}^{-1} + c$$ Equation 5

$$V_{basc} = (Q_{ref} * \Delta t)/n \quad \text{Equation 6}$$

$$V_{basc} = b_0 + b_1 Q_{Basc} + b_2 Q_{Basc}^2 \quad \text{Equation 7}$$

Where: Q_{tip} is the flow rate measured by the TB; Q_{ref} is the reference flow rate; m is the slope of the linear regression curve; a and b are constants of the potential regression curve, in which a is the point of intercession when the simulated flow is equal to 1 and b is the curve slope; T is the time between dumps; Δt is the time interval (1, 2, 5, 10, 20 or 30 minutes); n is the number of dumps registered; c is the constant in the T vs. 1/Q curve indicating the time required for the cavity to leave the stable at one side, move and reach the stable point at the other side; and b_0, $b_1$and b_2 are constants of adjustment in the quadratic curve.

2.3 Statistical analysis

For the statistical validation of TB applicability for flow monitoring during the dynamic calibration tests, three statistical metrics were applied: coefficient of determination (R^2); percentage of bias (PBIAS); and Kling-Gupta efficiency (KGE) in a non-parametric form. The R^2 assesses the degree of collinearity between measured and reference flows, varying between 0 and 1. The closer to 1, the better the correlation between measured and reference data (Surfleet *et al.*, 2012). The PBIAS indicates the average tendency of the measured data to be larger or smaller than the ones observed (Gupta *et al.*, 1999). Positive PBIAS values indicate a measured underestimation of models and equipment representing reference data, while negative values mean an overestimation and, when equal to zero, a perfect correlation of the data. Finally, using the KGE metric in the non-parametric form was an option in an attempt to use a more robust function that would allow analysis through different aspects (BIAS, standard deviation and Pearson's correlation), as indicated by Pool *et al.* (2018). In this metric, the values vary between 0 and 1; the closer to 1, the better the statistical correlation between the measured and reference data.

In addition to using statistical metrics that help the description, spatialization and comparison of the simulated data, the Analysis of Variance (2-way ANOVA) and Pearson's correlation were applied to investigate the interference of the selected variables (tipping speed, cavity dimension and time interval) in the mean error between the observed and the simulated flow rates at TBs of different sizes.

3 Results and discussion

3.1 Static calibration

The mean and standard deviations of measures at each cavity and global analysis (both cavities) were obtained during the static calibration of each TB (Table 1). After performing the procedure and calculating the mean of the measurements, it was found that the nominal volumes were 11.63 mL, 64.16 mL, 139.86 mL and 660.95 mL for TB1, TB2, TB3 and TB4, respectively. Thus, these will be the values used to identify the volumes and flows measured during the dynamic calibration, calculated by its product with the registered number of tips.

The standard deviation (SD), when expressed in absolute values (mL), has a positive correlation with the size of the equipment, ranging from 0.44 to 12.21 in the TB1 and TB4, respectively. However, when expressed in percentage values, TB1 has a higher SD (3.82%) than TB4 (1.85%). Among the various factors that could lead to such a result, it is believed that it may be associated with the cavity small water storage capacity and great sensitivity of TB1. Although care has been taken to carry out the calibration through the slow dripping of water, the kinetic effect added to the drop volume promotes oscillations between the replicates (Iida *et al.*, 2012). As the cavity size increases, this effect is smoothed out and, therefore the SD decreases.

INSERT TABLE 1

3.2 Dynamic calibration

Table 2 give the mean, maximum and minimum errors, standard deviation, PBIAS and KGE in TBs under different time intervals. The number of tips registered at each of the flow sampling points in each TB is available as Supplementary Material (A). In both TBs, it is observed that time interval plays a fundamental role in the calibration process. As expected, in shorter intervals, there is a smaller number of data records, and thus the SD is greater than when using longer intervals, such as 30 minutes. This point will be better discussed in the following sections.

Considering the ten flow rates sampled during the dynamic calibration process, both data from all TBs registered positive PBIAS (underestimation of reference flow). The highest PBIAS (13.6%) was observed in TB4 under a nominal volume of 660.95 mL, followed by TB3 (9.3%), TB2 (5.7%) and TB1 (3.3%), which have nominal volumes of 139.86 mL, 64.16 mL and 11.63 mL, respectively. The errors occurred under a range of low and high runoff intensities and TBs could still operate adequately and even before applying calibration curves, the proposed monitoring equipment can adequately measure the water flow (KGE> 0.86).

Given the flow ranges analyzed in both TBs, a greater standard deviation of errors was recorded at low flows, reducing as the flow increased, as well as in rainfall gauges (Shedekar *et al.*, 2016). Among all the designed equipment, TB1 had the highest SD, especially under a small-time interval (1 minute). It is important to cite that the reed switch in TB1 is located below the central axis, between cavities 1 and 2, recording one electrical signal every two tipping points. This contributes to the error being greater in this equipment when compared to the others, which have a record at each tip. Under an increasing time interval (30 minutes) and flow rate, the influence of this limitation reduces and, consequently, the SD is smaller. Thus, monitoring short precipitation events has a greater associated error than those with longer duration. Similarly, at higher intensities, there is a reduction in SD while there is a higher mean error. At any measured flow rates, there was an overestimation of the reference flows. The behavior of TBs under flow rates at other time intervals analyzed can be found in Supplemental Material (B).

Sun *et al.* (2014) designed and calibrated TBs with a nominal resolution of 2.5 liters and identified the same error pattern: high errors under low and high flow rates. It is believed that under low flow rates, the surface tension of the water influences the displacement along the surface of the cavity (Sun *et al.*, 2014), while at high flow rates, the slow and subtle shift, ideal in the gravity center, is affected by the rapid entry of water under turbulent flow (Iida *et al.*, 2012). Another error source comes from the water left in the cavities after one replicate test ending, which is not sufficient to tip (Nehls *et al.*, 2011). This volume was not removed from the cavities between calibration tests since we wanted to estimate the errors that would occur during in-situ monitoring and identify the best calibration model to reduce those errors.

INSERT TABLE 2

3.2.1 Linear regression

As can be seen, in all TBs there was an underestimation of flow rates as the angular coefficients (m) obtained are lower than 1 (Figure 3). From the data given in Figure 4 and Figure 5, it can be seen that for both TBs, the time interval is a relevant variable that influences the coefficient of determination (R^2). However, there is no direct correlation between this variable and the angular coefficient (m) of the regression curve. In this section, we will discuss data obtained under 30 minutes of time interval, but you can find data regarding other time intervals in Supplemental Material (C).

Using linear regression is a satisfactory option for all TBs analyzed, as shown by the statistical metrics used: R^2 (> 0.99) and KGE (> 0.85). As given in the previous session, the PBIAS registered in TBs has a direct correlation with the nominal volume of its cavities. Likewise, the slope of the linear fit curve (0.967, 0.942, 0.895 and 0.852) follows the same trend for TB1, TB2, TB3 and TB4, respectively. After implementing the curves, the fitting curve was ideal in TB1 (PBIAS = 0), while underestimation still occurred in TB2 (0.079%), TB3 (1.189%) and TB4 (1.397%). Similarly, the KGE index, considering PBIAS in its calculation, has an inversely proportional (Pearson correlation of -0.905) and not significant (p-value of 0.095) behavior for TB1, TB2, TB3 and TB4: 0.967, 0.942, 0.896 and 0.852.

The results are similar to those obtained by Khan and Ong (1997), Yahaya *et al.* (2009) and Sun *et al.* (2014) after applying the linear curves to reduce errors during the calibration of TBs of different sizes. Khan and Ong (1997) carried out the calibration process of a TB with a volumetric capacity of 3 liters and obtained R^2 equal to 0.99 and a residual overestimation error of 2%. Similarly, Yahaya *et al.* (2009) obtained a good coefficient of determination of 0.99 in the calibration of a 0.14 liter of volumetric capacity, which had an average error of 0.74%. Finally, Sun et al. (2014) calibrated a TB with an NV of 2.5 liters, finding a good linear correlation (R^2 equals to 0.99) between reference and measured flows and low mean error (2.1%). It is important to note that the NV of TB4 (660.95 mL) is greater than all of those previously mentioned, which would then be expected to have a high error, however, its average error is lower (1.397%), proving its efficiency.

INSERT FIGURE 3

INSERT FIGURE 4

INSERT FIGURE 5

3.2.2 Potential regression

The potential curve (Figure 6) is a satisfactory option for all of the TBs, as shown by the statistical metrics used: R^2 (> 0.99), KGE (> 0.85) and mean residual error (<0.2). As recorded in the linear regression curve, the KGE has an inversely proportional correlation (Pearson's correlation of -0.906) with NV, although it was not statistically significant (p-value of 0.094). As for PBIAS, there is an overestimation of the residual error of -0.358%, -0.251%, -0.021% and -0.183% in TB1, TB2, TB3 and TB4, respectively. Unlike what was observed in the linear curve, there is no clear correlation between PBIAS and NV. Barfield and Hirschi (1986) found overestimated errors between 1.62% and 1.90% in four TBs with NV ranging between 356mL and 1284mL while applying a potential calibration curve.

INSERT FIGURE 6

3.2.3 T vs. 1/Q regression

In Figure 7, it can be observed that the method has limitations under two situations: low flow rates and short time intervals (Shedekar *et al.*, 2016). Although the flow sampling points are uniformly distributed throughout the sampling range due to the mathematical formulation of the method, there is a concentration of sampling points at the bottom curve, while just few sampling points contribute to adjusting the curve in the upper portion.

Besides the low reliability at low flow rates, this method has good metrics statistics: R^2 (> 0.99), KGE (> 0.85) and mean residual error (<3%). By implementing this curve, there is an overestimation of residual error in TB1 (5.5%), TB2 (4.55%) and TB4 (1.68%), while there is an underestimation (0.36%) in TB3.

INSERT FIGURE 7

3.2.4 Quadratic regression

Based on R^2, this method works well (R^2 higher than 0.6) for TB1 (0.721), TB3 (0.826) and TB4 (0.615). It is important to note that although the KGE presented satisfactory values, the sampling points are completely dispersed (R^2 of 0.0255) along the regression curve (Figure 8) in the TB2. The statistical discrepancies observed are associated with the KGE mathematical formulations involved, and thus emphasize the importance of using different metrics in the calibration process.

By implementing this curve, there is an underestimation in the residual error in TB1 (0.747%), TB3 (3.114%) and TB4 (0.134%), while there is an overestimation (-0.129%) in TB2. Although we have not found the quadratic to be a satisfactory method for the TBs calibration, Somavilla *et al.* (2019) obtained a good statistical correlation (R^2 of 0.99 and NSE of 0.997) and low underestimation (2.27%). Similarly, Shimizu *et al.* (2018) were successful in eliminating the 2–3% errors in steamflow measurements. However, in both cases, TBs have a low volumetric capacity.

INSERT FIGURE 8

3.2.5 Determining the best calibration curve

After applying different adjustment curve methods, it was defined which curve has the best fit. Keeping in mind the importance of standardizing methodologies for monitoring runoff in the study area, the linear regression has the best statistical metric values and it also has greater simplicity and confidence in terms of extrapolation. Table 3 shows the calibration equations to obtain the flow measured by the TBs based on the number of tips counted by the datalogger. Note that the multiplier number is greater than the nominal resolution found during static calibration due to the underestimate error sources presented before, such as water tension, kinetic effects, water left in cavities and spills.

INSERT TABLE 3

3.3 Causality tests

3.3.1 Tipping velocity

Through the Pearson correlation, the possibility of a correlation between the tipping speed and the percentage error recorded in the different TB sizes was investigated. It can be observed that the TBs have discordant behaviors: TB3 presented a directly proportional (0.852) significant correlation (p-value of 0.02), while TB1 and TB2 had negative (-0.579 and -0.008) and TB4 positive (0.707) correlations, but both not significant (p-value > 0.05). Considering a joint analysis (join data from all TBs), we found a positive correlation (0.134) and also significant one (p-value equals to 0.038).

When plotting the mean error according to the number of dumps registered in the TBs (Figure 9), a similar pattern was identified in the error curve behavior, which can be summarized in three zones: Zone I occurs at low tipping speeds and results in high errors (underestimation); zone II occurs at average tipping speeds, characterized by a decay of underestimation and thus, it is considered as the optimal range of operation; finally, in zone III at a high tipping speed, the percentage error re-rises. While calibrating a TB with a 0.14 liter NV in Nigeria, Yahaya *et al.* (2009) also agree that the runoff intensity and the tipping rate intefere greatly in the errors and, finally, after plotting the efficiency versus the runoff errors, found a parable trend in the data collected, implying the existence of these three zones explained here.

In general, there is a tendency to underestimate the flow measured by TBs under high intensities of runoff (Iida *et al.*, 2012; Somavilla *et al.*, 2019) and rainfall (Shedekar *et al.*, 2016; Sypka, 2019). This phenomenon can be attributed to the volume of water lost while cavities switches (Shedekar *et al.*, 2016). As the water enters a constant flow, when reaching the nominal volume, the cavity starts the switching but the water continues to fall on the cavity that already reached its NV. Thus, there is a small time interval for the filled cavity to move and water to begin to fall into the second cavity. This delay was also found by Langhans *et al.* (2019) in TBs of different NVs (0.1-2 l). Considering that this displacement interval is constant, under increased flow of water, the greater the underestimation errors recorded (Edwards *et al.*, 1974).

INSERT FIGURE 9

3.3.2 Time interval of data collection

The specification of a time interval for the hydrological data acquisition commonly depends on the available data storage capacity and time interval of other installed equipment. Using shorter time intervals, the greater the possibilities of recording extreme events, the better the monitoring of natural phenomena (Shedekar *et al.*, 2016).

Different data recording intervals were used in the sampling during the dynamic calibration (1, 2, 5, 10, 20, and 30 minutes). Thus, to investigate the interference that this variable has in the mean percentage error, the graphs of the main effects were drawn up (available in Supplemental Material D). Through the graphs, it can be observed that there is a positive correlation in the TB1, TB3 and TB4, while the TB2 has a negative correlation. In order to validate the correlation identified through the graphs of the main effects, the Pearson's correlation was calculated between the time interval and the mean errors. It can be noted that there is no strong correlation (Pearson's correlation less than 0.2) between the variables analyzed, however, due to the p-value being above the limit (0.05), the null hypothesis considered cannot be rejected.

TB1, TB2 and TB4 had a greater error variation at different sampling time intervals, while TB3 had a smoother variation, considering different sampling time intervals (1, 2, 5, 10, 20 and 30 minutes) (Figure 10). As specially observed in TB1 and TB2, shorter intervals result in greater errors (Habib *et al.*, 2001; Shedekar *et al.*, 2016). At low flow rates, when the time interval required to reach the nominal volume is higher than the time interval, two or more time intervals are required to register a tipping. In the first-time interval, there is no record of tipping, as it was only recorded in the second.

INSERT FIGURE 10

Ciach (2003) and Costello and Williams (1991) also found that the sampling time interval is a significant variable in hydrological studies and indicate the use of the tipping interval instead of defined sampling times. However, it was not possible to adopt such a consideration due to limitations in the datalogger used (*Campbell Scientific Inc* CR10), which had the capacity to record data with a minimum interval of one minute. Considering this, it is recommended that further studies be carried out on the errors in TBs by identifying the time between the emptying of one cavity and the beginning of filling the other one.

3.3.3 Tipping bucket volumetric capacity

The third variable investigated with the potential to influence the mean errors was the TBs´ volumetric capacity. Through the Pearson correlation test, we found a statistically significant (p-value $\leq$ 0.05) positive correlation (0.369), indicating that mean errors are directly influenced by NV. The obtained data reinforces the importance of the adequate sizing of TBs, so that it is not under or over-sized.

The results obtained follow the consideration of Shedekar *et al.* (2016) and Somavilla *et al.* (2019) that the storage capacity of TBs is an important source of errors. As previously mentioned, the underestimation is possibly due to the volume of water lost during the time interval of cavity switching. The volume of water lost is a fraction of the volume stored in the cavities, the greater the storage capacity, the lower the sensitivity of the equipment to this small fraction of volume that is not monitored, and thus the greater the associated errors (Somavilla *et al.*, 2019).

Finally, another important source of error is the volume of residual water retained in the cavities which is not enough for tipping. For example, TB1 has an NV of 11.63 mL, while TB4 has a 660.95 mL. TB1 has a higher volumetric resolution than TB4 and, consequently, a smaller residual volume that can be lost by evaporation.

3.3.4 Joint analysis

As previously presented, the percentage error is influenced by the variables´ tipping speed, cavity capacity and time interval. In order to investigate which variable analyzed has the greatest contribution to the mean error, we applied the Multi-factor Analysis of Variance. We found out that the cavity size has a greater influence (F-value of 16.12) in mean errors than tipping speed (F-value of 11.34). The null hypothesis of non-correlation between the variables´ time interval and mean errors cannot be rejected, due to the p-value above the imposed limit of 0.05.

3.4 Factors affecting calibration and errors

The operation principle of TBs used for runoff and rain measurements is the same, thus the errors to which they are susceptible are similar. Errors can be grouped into two categories: systematic/mechanical and random (Shedekar *et al.,* 2016). Systematic errors are due to the operation, construction material and design of the equipment, thus they are more predictable and easier to minimize. Random errors, however, are not predicted and occur from unusual operations during in-situ measurements.

In addition to some main examples of systemic and random errors given in Table 4, there are those already mentioned and discussed previously (tipping speed and time interval) and errors inherent in any laboratory measurement, in this case: uncertainties in the measurements of the nominal volume and reference flow rate during static and dynamic calibration, respectively. Despite the inherent error sources, the use of tipping bucket for runoff measurement is a potential instrument for a better understanding in the hydrology field. It is even more applicable in developing countries, such as Brazil, which most of the time have limited funding for acquisition of high-tech monitoring equipment.

INSERT TABLE 4

4 CONCLUSIONS

A tipping bucket flow meter is a potential instrument in in-situ monitoring. However, its inherent errors may limit and create doubts about its data reliability. Thus, in this paper, we performed static and dynamic calibration on four large TBs, applied different regression curves and investigated the error sources in order to minimize them. Through static calibration, we found out that the equipment had a volumetric capacity of 11.63 mL (TB1), 64.16 mL (TB2), 139.86 mL (TB3) and 660.95mL (TB4). Afterwards, TBs were tested under different flow rates (dynamic calibration) and time intervals. Considering a 30-minute time interval, an underestimation of flows at different levels was identified according to the size of the cavity: TB1 (3.31%), TB2 (5.75%), TB3 (9.33%) and TB4 (13.57%).

We discovered a high underestimated error at low flow rates, indicating that the best operating range of the equipment is under medium flow rates. Even not dimensioned to operate at high intensities, the equipment was tested in the laboratory under different low and high flow intensities and had a satisfactory performance.

After performing the dynamic tests, four calibration equations were tested: linear, potential, T vs. 1/Q and quadratic. Among the alternatives, linear regression showed the best correlation (above 99%) with the monitored data. Using this method, the mean error will be reduced to -1.35% (TB1), 0.04% (TB2), 3.18% (TB3) and 3.73% (TB4).

Once the occurrence of systematic errors was verified, it was investigated how the different variables (tipping speed, tipping bucket volumetric capacity and time interval of data collection) influenced the errors. By plotting mean errors and tipping rates, a behavior pattern in the error curves was identified: Zone I at low tipping speeds and results in high percentage errors (underestimation); zone II at average tipping speeds, characterized by a decay of underestimation and thus, it is considered as the optimal range of operation; and zone III at high tipping speed, the percentage error re-rises. Investigating the second variable, we found that mean errors are directly influenced by the cavity volumetric capacity. The last variable, time interval of data collection, barely interfered in the data sampled. Finally, considering all three error sources, the cavity size is the most important one.

Throughout the research, we aimed to minimize the errors inherent to the large tipping buckets and encourages various applications (steam, runoff, percolation, etc.), increasing the in- situ hydrological data collection, which is still very scarce, mainly in developing countries.

5 DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

REFERENCES

Anache JAA, Flanagan DC, Srivastava A, Wendland EC. 2018. Land use and climate change impacts on runoff and soil erosion at the hillslope scale in the Brazilian Cerrado. *Science of the Total Environment* 622–623: 140–151 DOI: 10.1016/j.scitotenv.2017.11.257

Anache JAA, Wendland E, Rosalem LMP, Youlton C, Oliveira PTS. 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

Barfield BJ, Hirschi MC. 1986. Tipping Bucket Flow Measurements on Erosion Plots. *Transactions of the ASAE* 29 (6): 1600–1604 DOI: 10.13031/2013.30360

Calder IR, Kidd CHR. 1978. A note on the dynamic calibration of tipping-bucket gauges. *Journal of Hydrology* 89: 383–386 DOI: 10.1016/0022-1694(78)90013-6

Chanapathi T, Thatikonda S. 2020. Investigating the impact of climate and land-use land cover changes on hydrological predictions over the Krishna river basin under present and future scenarios. *Science of the Total Environment* 721: 137736 DOI: 10.1016/j.scitotenv.2020.137736

Chow TL. 1976. Overland Flow and Subsurface Stormflow Studies. *Canadian Journal of Soil Science* 56: 197–202

Ciach GJ. 2003. Local random errors in tipping-bucket rain gauge measurements. *Journal of Atmospheric and Oceanic Technology* 20 (5): 752–759 DOI: 10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2

Corona R, Wilson T, D'Adderio LP, Porcù F, Montaldo N, Albertson J. 2013. On the Estimation of Surface Runoff through a New Plot Scale Rainfall Simulator in Sardinia, Italy. *Procedia Environmental Sciences* 19: 875–884 DOI: 10.1016/J.PROENV.2013.06.097

Costello TA, Williams HJ. 1991. Short duration rainfall intensity measured using calibrated time-of-tip data from a tipping bucket raingage. *Agricultural and Forest Meteorology* 57 (1–3): 147–155 DOI: 10.1016/0168-1923(91)90083-3

Edwards IJ, Jackon WD, Fleming PM. 1974. Tipping bucket gauges for measuring run-off from experimental plots. *Agricultural Meteorology* 13 (2): 189–201

Egorov Y V., Bobkov A V., Esafova EN, Fless AD. 2015. Installation for determining the solid and liquid runoffs during storm erosion. *Eurasian Soil Science* 48 (2): 218–222 DOI: 10.1134/s1064229315020039

Elder K, Marshall HP, Elder L, Starr B, Karlson A, Robertson J, Forest US, Rmrs S, Collins F. 2014. Design and Installation of a Tipping Bucket Snow Lysimeter. In *International Snow Science Workshop 2014 Proceedings, Banff Canada*Montana State University Libraryskip navigation: Banff, Canada; 817–824. Available at: http://arc.lib.montana.edu/snow-science/item/2157

Gupta HV, Sorooshian S, Yapo PO. 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *Journal of Hydrologic Engineering* 4 (2): 135–143 DOI: 10.1061/(ASCE)1084-0699(1999)4:2(135)

Habib E, Krajewski WF, Kruger A. 2001. Sampling Errors of Tipping-Bucket Rain Gauge Measurements. *Journal of Hydrologic Engineering* 6 (2): 159–166 DOI: 10.1061/(asce)1084-0699(2001)6:2(159)

Hollis GE, Ovenden JC. 1987. A tipping bucket flowmeter for roadside gully runoff. *Hydrological Processes* 1: 391–396 DOI: 10.1002/hyp.3360010407

Humphrey MD, Istok JD, Lee JY, Hevesi JA, Flint AL. 1997. A New Method for Automated Dynamic Calibration of Tipping-Bucket Rain Gauges. *Journal of Atmospheric and Oceanic Technology* 14: 1513–1519

Iida S, Shimizu T, Kabeya N, Nobuhiro T, Tamai K, Shimizu A, Ito E, Ohnuki Y, Abe T, Tsuboyama Y, et al. 2012. Calibration of tipping-bucket flow meters and rain gauges to measure gross rainfall, throughfall, and stemflow applied to data from a Japanese temperate coniferous forest and a Cambodian tropical deciduous forest. *Hydrological Processes* 26: 2445–2454 DOI: 10.1002/hyp.9462

Johnston CN. 1942. Tilt buckets for measuring runoff and erosion. *Agriculture engineering* 23 (5): 161–162

Kanzari S, Ben Nouna B, Ben Mariem S, Rezig M. 2018. Hydrus-1D model calibration and validation in various field conditions for simulating water flow and salts transport in a semi-arid region of Tunisia. *Sustainable Environment Research* 28 (6): 350–356 DOI: 10.1016/j.serj.2018.10.001

Khan AAH, Ong CK. 1997. Design and calibration of tipping bucket system for field runoff and sediment quantification. *Journal of Soil and Water Conservation* 52 (6): 437–443

Kifle Arsiso B, Mengistu Tsidu G, Stoffberg GH, Tadesse T. 2017. Climate change and population growth impacts on surface water supply and demand of Addis Ababa, Ethiopia. *Climate Risk Management* 18: 21–33 DOI: 10.1016/j.crm.2017.08.004

Kim HJ, Sidle RC, Moore RD. 2005. Shallow lateral flow from a forested hillslope: Influence of antecedent wetness. *CATENA* 60 (3): 293–306 DOI: 10.1016/J.CATENA.2004.12.005

Klik A, Sokol W, Steindl F. 2004. Automated erosion wheel: A new measuring device for field erosion plots. *Journal of Soil and Water Conservation* 59 (3): 116–121

Lamb KJ, MacQuarrie KTB, Butler KE, Danielescu S, Mott E, Grimmet M, Zebarth BJ. 2019. Hydrogeophysical monitoring reveals primarily vertical movement of an applied tracer across a shallow, sloping low-permeability till interface: Implications for agricultural nitrate transport. *Journal of Hydrology* 573: 616–630 DOI: 10.1016/J.JHYDROL.2019.03.075

Langhans C, Diels J, Clymans W, Van den Putte A, Govers G. 2019. Scale effects of runoff generation under reduced and conventional tillage. *Catena* 176: 1–13 DOI: 10.1016/j.catena.2018.12.031

Mahlknecht J, González-Bravo R, Loge FJ. 2020. Water-energy-food security: A Nexus perspective of the current situation in Latin America and the Caribbean. *Energy* 194: 116824 DOI: 10.1016/j.energy.2019.116824

Mello K De, Taniwaki RH, Paula FR de, Valente RA, Randhir TO, Macedo DR, Leal CG, Rodrigues CB, Hughes RM. 2020. Multiscale land use impacts on water quality: Assessment, planning, and future perspectives in Brazil. *Journal of Environmental Management* 270 (June): 110879 DOI: 10.1016/j.jenvman.2020.110879

Nebol'sin SI. 1928. Elementary Surface Runoff. Moscow.

Nehls T, Rim YN, Wessolek G. 2011. Hydrology and Earth System Sciences Technical note on measuring run-off dynamics from pavements using a new device: the weighable tipping bucket. *Hydrology and Earth System Sciences* 15: 1379–1386 DOI: 10.5194/hess-15-1379-2011

Nóbrega RLB, Guzha AC, Torres GN, Kovacs K, Lamparter G, Amorim RSS, Couto E, 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

Perales-Momparler S, Andrés-Doménech I, Hernández-Crespo C, Vallés-Morán F, Martín M, Escuder-Bueno I, Andreu J. 2017. The role of monitoring sustainable drainage systems for promoting transition towards regenerative urban built environments: a case study in the Valencian region, Spain. *Journal of Cleaner Production* 163: S113–S124 DOI: 10.1016/J.JCLEPRO.2016.05.153

Peyrard X, Liger L, Guillemain C, Gouy V. 2016. A trench study to assess transfer of pesticides in subsurface lateral flow for a soil with contrasting texture on a sloping vineyard in Beaujolais. *Environmental Science and Pollution Research* 23 (1): 14–22 DOI: 10.1007/s11356-015-4917-5

Pool S, Vis M, Seibert J. 2018. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal* 63 (13–14): 1941–1953 DOI: 10.1080/02626667.2018.1552002

Rocha J, Carvalho-Santos C, Diogo P, Beça P, Keizer JJ, Nunes JP. 2020. Impacts of climate change on reservoir water availability, quality and irrigation needs in a water scarce Mediterranean region (southern Portugal). *Science of the Total Environment* 736: 139477 DOI: 10.1016/j.scitotenv.2020.139477

Rosalem LMP, Anache JAA, Wendland E. 2018. Determining forest litter interception in an area of the Cerrado sensu stricto. *Revista Brasileira de Recursos Hidricos* 23 (23) DOI: 10.1590/2318-0331.231820170146

Shedekar VS, King KW, Fausey NR, Soboyejo ABO, Harmel RD, Brown LC. 2016. Assessment of measurement errors and dynamic calibration methods for three different tipping bucket rain gauges. *Atmospheric Research* 178–179: 445–458 DOI: 10.1016/j.atmosres.2016.04.016

Shimizu T, Kobayashi M, Iida S, Levia DF. 2018. A generalized correction equation for large tipping-bucket flow meters for use in hydrological applications. *Journal of Hydrology* 563 (January): 1051–1056 DOI: 10.1016/j.jhydrol.2018.06.036

Shiraki K, Yamato T. 2004. Compensation of Tipping-Bucket Flow Meters. *Journal of Japan society of hydrology & water resources* 17: 159–162 DOI: 10.3178/jjshwr.17.159

Somavilla A, Gubiani PI, Zwirtz AL. 2019. Tipping bucket prototype for automatic quantification of surface runoff rate in plots. *Revista Brasileira de Ciencia do Solo* 43: 1–7 DOI: 10.1590/18069657rbcs20180096

Sun T, Cruse RM, Chen Q, Li H, Song C, Zhang X. 2014. Design and initial evaluation of a portable in situ runoff and sediment monitoring device. *Journal of Hydrology* 519: 1141–1148 DOI: 10.1016/j.jhydrol.2014.08.048

Surfleet CG, Tullos D, Chang H, Jung IW. 2012. Selection of hydrologic modeling approaches for climate change assessment: A comparison of model scale and structures. *Journal of Hydrology* 464–465: 233–248 DOI: 10.1016/j.jhydrol.2012.07.012

Sypka P. 2019. Dynamic real-time volumetric correction for tipping-bucket rain gauges. *Agricultural and Forest Meteorology* 271 (July 2018): 158–167 DOI: 10.1016/j.agrformet.2019.02.044

Takahashi M, Giambelluca TW, Mudd RG, DeLay JK, Nullet MA, Asner GP. 2010. Rainfall partitioning and cloud water interception in native forest and invaded forest in Hawai'i Volcanoes National Park. *Hydrological Processes* 25 (3): 448–464 DOI: 10.1002/hyp.7797

Wang S, Fu Z, Chen H, Nie Y, Xu Q. 2020. Mechanisms of surface and subsurface runoff generation in subtropical soil-epikarst systems: Implications of rainfall simulation experiments on karst slope. *Journal of Hydrology* 580: 124370 DOI: 10.1016/J.JHYDROL.2019.124370

Whipkey RZ. 1965. Subsurface stormflow from forest slopes. *Hydrological Sciences Journal*

10 (2): 74–85 DOI: 10.1080/02626666509493392

Yahaya O, Olufayo AA, Akinro AO, Mogaji Kehinde O. 2009. Development and calibration of an automatic Runoff-meter. *Journal of Engineering and Applied Sciences* 4 (9): 8–16

Yang YCE, Son K, Hung F, Tidwell V. 2020. Impact of climate change on adaptive management decisions in the face of water scarcity. *Journal of Hydrology* 588: 125015 DOI: 10.1016/j.jhydrol.2020.125015

Youlton C, Bragion AP, Wendland E. 2016a. Experimental evaluation of sediment yield in the first year after replacement of pastures by sugarcane. *Cien. Inv. Agr* 43 (3): 374–383 DOI: 10.4067/s0718-16202016000300004

Youlton C, Wendland E, Anache JAA, Poblete-Echeverría C, Dabney S. 2016b. Changes in erosion and runoff due to replacement of pasture land with sugarcane crops. *Sustainability (Switzerland)* 8 (7): 1–12 DOI: 10.3390/su8070685

Zabret K, Rakovec J, Šraj M. 2018. Influence of meteorological variables on rainfall partitioning for deciduous and coniferous tree species in urban area. *Journal of Hydrology* 558: 29–41 DOI: 10.1016/J.JHYDROL.2018.01.025