DOI:10.28974/idojaras.2022.2.7

# IDŐJÁRÁS

*Quarterly Journal of the Hungarian Meteorological Service*
*Vol. 126, No. 2, April – June, 2022, pp. 285–295*

## Short Contribution

# On the correction of multiple minute sampling rainfall data of tipping bucket rainfall recorders

**Tibor Rácz**

*Doctoral School of Environmental Sciences*
*at the Hungarian University of Agriculture and Life Sciences*
*Páter Károly utca 1, Gödöllő, 2100, Hungary*

Author E-mail: *raczt167@gmail.com*

*(Manuscript received in final form February 8, 2021)*

**Abstract**— In the last decades of the 1900s, the tipping bucket rainfall gauges (TBG) were used to record the sub-daily rainfall data. In the first period of the rainfall data recording, as a result of the lack of efficient data storage and data transmission, the sampling period of the TBG devices was chosen in a magnitude of 10–20 minutes. Consequently, there are historical datasets characterized by several minutes long sampling periods. Since the turn of the 2000s, the data handling has been revolutionized; the sampling period has diminished to one minute. There is a systematic error of the TBG technique which has been investigated since the middle of the 1900s. Between 2004 and 2008, a comprehensive research was performed to determine the correction equation for several TBG devices. These results can be utilized for the short sampling period measurements (one minute sampling), but for longer sampling period data, further corrections are needed. In this paper, a supplementary correction is presented. On the base of the mathematical determination of the correction factor, simple estimation will be proposed to be able to execute the necessary correction. After the presentation of the correction factor, a general correction factor is proposed for larger geographical regions and wide time span of the measurements. The revision of the historical rainfall data recorded by TBG devices can be important in several issues, such as the re-evaluation of intensity-duration-frequency (IDF) curves, and in other fields.

*Key-words:* historical rainfall data, tipping bucket gauge, rainfall recorder, rainfall intensity, data correction

## *1. Introduction*

The measurement of the rainfall intensities has great importance in providing key data of engineering hydrology for the design works of drainage systems or flood protection interventions. Although the measurement of the rainfall intensity has a 300-year-long history (*Kurytka*, 1953), the importance of the rainfall intensity was recognized in its significance only in the middle of the 19th century, after Mulvany has worked out the rational method (*Mulvany*, 1851). First, the rainfall intensity measurement was performed by rainfall writers, detecting the changing water level in the tank of the device. Their dominance was evident till the 1970s. Because of the simpler electronic data recording and ready-to-process data format, the water level writing gauges got behind the simpler, smaller TBG devices in the practice. These devices have a long history; Sir Christopher Wren built the first tipping bucket gauge in the second half of the 1600s, in an early phase of the development of modern rainfall measurement devices. In the following centuries, there were several arrangements of these kinds of gauges. The appearance of the electronic data registration and the quite simple processing option of the data performed by TBG units have made the technology wide-spreading.

As every measurement technique, the TBG has its systematic error. Beyond the usual sources of error, such as the wind-caused uncertainty of the data, there are structural sources of errors, the so-called local random errors (*Habib et al.*, 2013). The nature of these errors are explained comprehensively in the related papers, and there are several correction methods to diminish or exclude it (*Marsalek,* 1981; *Adami* and *Da Deppo* 1985; *Niemczynovicz,* 1986; *Habib et al.,* 2001; *Luyckx* and *Berlamont,* 2001; *Frankhauser,* 1997; *Vuerich et al.,* 2009; *Duchon* and *Biddle,* 2010). These solutions are fundamental for the short sampling period (characteristically one-minute) measurements. However, for an old time series having longer sampling period, $n$ times longer than the one-minute unit interval (so $n$-minute-long), a further correction method is required. In this paper, a proposal for a supplementary correction is presented, for 5–10 minutes or longer intervals data.

## *2. Methods*

The proposed correction for longer sampling periods is based on the results of the measurements and elaborated correction formulae of Vuerich and his colleagues (*Vuerich et al.,* 2009), using power function between the measured and reference data, as

$$i_c = a \cdot i_m^b, \tag{1}$$

where $i_c$ is the corrected value of the rainfall intensity (or reference intensity during the calibration process), $i_m$ is the measured value of the rainfall intensity, and $a, b$ are parameters related to the TBG rainfall gauge.

The rainfall intensity is the function of time and space, and it is a volumetric flux, so the $V(t)$ volume flows through an $A$ unit surface in a unit time interval, $i = i(t, x, y)$. The surface can be written as $A = A(x, y)$ and $dA = dx \cdot dy$, so the intensity can be determined as

$$i(t, x, y) = \frac{d}{dA}\frac{dV(t)}{dt}. \tag{2}$$

The volume of rainfall onto the unit surface in a $t = [T_1, T_2]$ interval can be expressed as

$$V(t) = \iint i(t)dAdt = \int_{T_1}^{T_2} i(t)dt, \tag{3}$$

but the unit of the rainfall intensity conserves the [volume/(area·time)] character.

In the practice, the measurement of the rainfall intensity by the TBG device occurs in finite time units (or sampling periods), in the last decades in one minute. The TBG device counts the number of rainfall volumes equal to the volume of the bucket in one sampling period. Assuming a longer measuring period of several minutes, where the $t$-long period is $n$ times longer than a supposed unit interval (so the measuring period is $n$ units long), the rainfall volume on a unit area can be calculated as the sum of the unit rainfall volumes of the unit interval, and it can be expressed with the unique intensities of the unit intervals too, so

$$V = \sum_{j=1}^{n} V_j = \sum_{j=1}^{n} \left(i_j \cdot \frac{t}{n}\right), \tag{4}$$

where $n$ is the number of the sub-intervals when the $t = [T_1, T_2]$ interval is divided to shorter intervals with equal lengths, and $i_j$ is the rainfall intensity of the given $\frac{t}{n}$-long interval.

Since the $i_{m,t}$ average intensity of the $t$-long interval can be expressed with the intensities of the $\frac{t}{n}$ sub-intervals as

$$i_{m,t} = \sum_{j=1}^{n} \frac{i_j}{n}, \tag{5}$$

the $V$ volume can be determined with the average intensity of the $t$ interval as well:

$$V = i_{m,t} \cdot t. \tag{6}$$

The volumes of the Eqs. (4) and (6) must be equal.

However, there is the fact that the TBG device measures rainfall together with its systematic error, so it must be corrected. Applying the correction formula Eq.(1) for Eqs.(4) and (6), the corrected volumes will be different in most of the cases. The question is on the one hand the measure of the difference, which depends on the $a$ and $b$ correction parameters and the $t$ length of the sampling period, and on the other hand, the demanded value of the supplementary correction.

For the empirical investigation of correction factors, the time series were chosen from the German Weather Service's (DWD's) one-minute rain depth database. (Source is: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/1_minute/precipitation/historical/2016/)

### *3. Results and discussion*

The measure of the supplementary correction will be shown for the $t$-long sampling period datasets, where in the practice the intensities of shorter intervals are not available, but now, a one-minute sampling period data will be used for the demonstration. For the first, a mathematical explanation will be shown, and then the supplementary correction will be presented using real data. Eq.(1) shows the method of the correction of a one-minute long sampling period rainfall data. If there is a $t$>1 minimal interval between the two measured data, the fallen rainfall volume (onto a unit area) can be calculated using the before mentioned two ways.

The first is a one-step correction on the base of the average rainfall intensity (see Eq.(6)), on the base of the available data of the $t$ sampling period; this is signed in the further part of this paper with the subscript "A". The corrected rainfall volume of the $t$ sampling period can be calculated as

$$V_{c,(A)} = a \cdot i_{m,t}^{b} \cdot t = a \cdot t \left( \sum_{j=1}^{n} \frac{i_j}{n} \right)^{b} = \frac{a \cdot t}{n^b} \left( \sum_{j=1}^{n} i_j \right)^{b}. \tag{7}$$

The other way (signed with "B") is the calculation of the sum of corrected $n$ sub-volumes (see Eq.(4)). This step cannot be done when data originate from a measurement with longer sampling period, since the sub-units are not available, but to point out the ratio of these volumes, its formulation must be done:

$$V_{c,(B)} = \sum_{j=1}^{n} \left( a \cdot i_j^{b} \cdot \frac{t}{n} \right) = \frac{a \cdot t}{n} \sum_{j=1}^{n} \left( i_j^{b} \right). \tag{8}$$

The intensity of the $j$th sub-interval can be expressed with a $c_j$ weighting factor, so

$$i_j = c_j \cdot i_{m,t} \, , \tag{9}$$

where $c_j$ is the a positive weighting factor for the $j$th sub-interval, and so with Eq.(5)

$$i_{m,t} = \frac{\sum_{j=1}^{n} i_j}{n} = \frac{\sum_{j=1}^{n} c_j \cdot i_{m,t}}{n} = i_{m,t} \cdot \frac{\sum_{j=1}^{n} c_j}{n}, \tag{10}$$

The consequence of Eq.(10) is that $\sum_{j=1}^{n} c_j = n$.

Eqs.(7) and (8) – using Eq.(9) – can be written as

$$V_{c,(A)} = \frac{a \cdot t}{n^b} \left(\sum_{j=1}^{n} i_j\right)^b = \frac{a \cdot t}{n^b} \left(\sum_{j=1}^{n} c_j \cdot i_{m,t}\right)^b = \frac{a \cdot t \cdot i_{m,t}{}^b}{n^b} \left(\sum_{j=1}^{n} c_j\right)^b, \tag{11}$$

$$V_{c,(B)} = \sum_{j=1}^{n} \left(i_{c,j} \cdot \frac{t}{n}\right) = \frac{t}{n} \sum_{j=1}^{n} \left(a \cdot i_{r,j}^b\right) = \frac{a \cdot t}{n} \sum_{j=1}^{n} \left((c_j \cdot i_{m,t})^b\right) = \frac{a \cdot t \cdot i_{m,t}{}^b}{n} \sum_{j=1}^{n} \left(c_j^b\right). \tag{12}$$

The ratio of the two volumes is, using that $\left(\sum_{j=1}^{n} c_j\right)^b = n^b$,

$$\frac{V_{c,(B)}}{V_{c,(A)}} = \frac{\sum_{j=1}^{n} (c_j^b)}{n} = CF_t. \tag{13}$$

This is the $CF_t$ correction factor which should be used to get the realistic $V_{c,(B)}$ multiplying the volume $V_{c,(A)}$ derivable from the available dataset, so the supplementary correction is

$$V_{c,(B)} = \frac{\sum_{j=1}^{n} (c_j^b)}{n} V_{c,(A)}. \tag{14}$$

The same is true for the intensities, as well, since the $V_c$ volumes are the $i_{m,t}$ multiplied with *t*, and expressing the average intensities in both sides of Eq.(14), *t* will fall out with a simplification, so the adjusted intensity (case "B")

$$i_{m,t,(B)} = \frac{\sum_{i=1}^{n} (c_i^b)}{n} \cdot i_{m,t,(A)} = \frac{\sum_{i=1}^{n} (c_i^b)}{n} \cdot a \cdot i_{m,t}^b. \tag{15}$$

If the intensity is not constant, i.e., when the $c_j$ values are different, then the value of Eq.(13) depends on the distribution of the $c_j^b$ weights (with a constant $b$). The number of this kinds of distributions can be infinite, however, the estimation of the $\sum_{j=1}^{n} (c_j^b)$ is possible, as it will be performed a little bit later. Before proposing an estimation method, a short discussion is needed to learn the main characteristics of the $CF_t$ values.

As Eq.(9) shows, if the $i_j$ intensity is constant in the $t$ interval, $c_j = 1$ for every $j$; in this case the $CF_t$ will get its lowest value, 1.

The weight numbers have an upper limit, as well. The highest value can be calculated if there is only one minute in the investigated interval when the rain depth differs from zero. In this case the only weight different from zero must be $n$, meanwhile the sum of the $b$th power of the weights is going to be $\sum_{j=1}^{n}\left(c_j^b\right) = n^b$ and the $CF_t$ value is to be $n^{b-1}$. So, the $CF_t$ values will be always in the $[1, n^{b-1}]$ bounded interval.

In the major part of the cases, the consecutive intensities are not really different, so their ratio varies not too much, and their values are characteristically near to 1. The highest values of the weights are at the transient period towards the very intensive rainfall, and after the peak, in the returning phase to the lower values. Interestingly, the periods of lower rainfall depths and intensities can be characterized with relatively higher $CF_t$ values. This is the consequence of the greater possibility of the varying of the rainfall intensities in these parts of the rainfall (it is simpler to change a lot from a low base value).

The frequency of the $CF_t$ values is strongly right tailed, and the mean and median of the $CF_t$ values are close to the lower limit.

From practical point of view, there is an issue with the low intensity data of the rainfall intensities. In the low intensity periods, there can be several 0 values, where the TBG device could have detect zero rain depths for several minutes, and there is one or some few measurements in the actual interval. This case demands a careful investigation, since if there was so slight rainfall that the tipping bucket could have been filled only after several minutes, the rain depth and intensity would show 0 values, despite of the continuous rain. In this case the $c_j$ values will show 0, however, their values should have been somewhere at 1, and there will be one weight with a high value, which should have been close to the other weights. This phenomenon is caused by the rainfall less intensive than the lowest measurable intensity in one minute. Intervals, which can be characterized by 70–80% 0 values of rainfall are not proposed to the calculation of $CF_t$ values, especially if the non-zero values are only the lowest measurable intensities. The main features of the $CF_t$ values with $b = 1.042$ for 5-minute intervals can be seen in *Fig. 1*. The source of data is the open database of 1-minute rainfall data of the DWD. The station is in Abtsgmünd-Untergröningen, and the rainfall was detected in June 11, 2018. (https://opendata.dwd.de/climate_environment/CDC/).

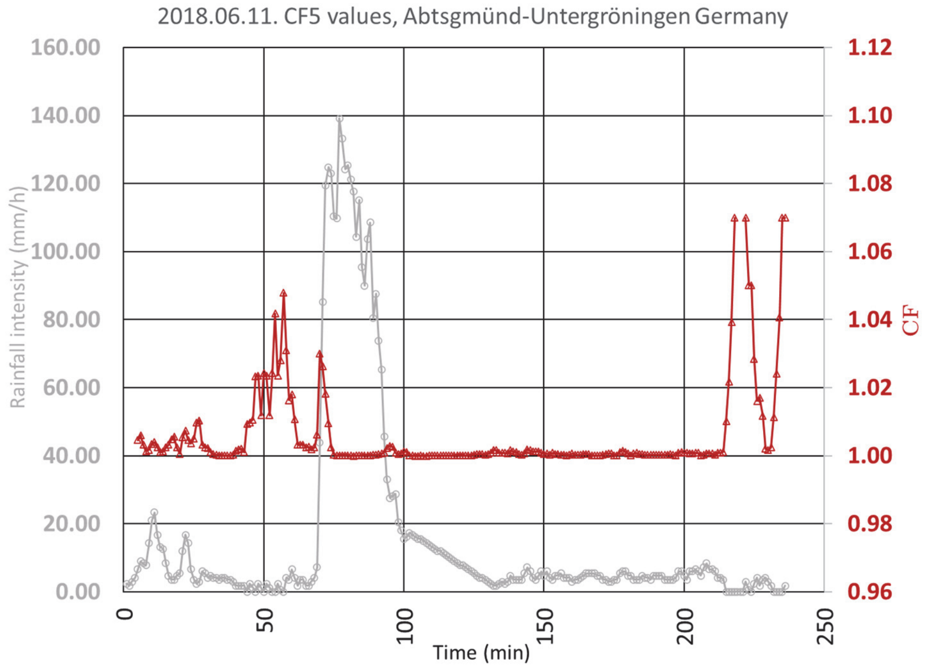


*Fig. 1*. The main characteristics of $CF_5$ values in the data of an intensive rainfall, $b$=1.042.

The above written characteristics of the $CF_5$ values can be observed in the *Fig. 1*. In the first 25 minutes of the rainfall, the one-minute sampling has resulted in continuously non-zero data, but with fluctuating intensity. The related $CF_5$ values show fluctuation, as well. Between the 44–57th and the 215–236th minutes, the rainfall intensity was so low that the bucket of the TBG device was not tipped for several minutes, and it resulted in several 0 mm rain depth and rainfall intensities, while despite of these, the rainfall was probably continuous; however, it is impossible to check anymore. Because of the nulls, several high $CF_5$ values can be observed, but probably these data are caused by the abovementioned issue of the low intensity that is below the intensity measurement resolution of the device; thus, these data are not confirmed. The most intensive part of the rainfall started in the 70th minute, where the fast rising of the rainfall intensity caused a high value in the $CF_5$ curve. In the highest range of the rainfall intensity – despite of a strong absolute fluctuation – the $CF_5$ values are low, and they rise again with the decreasing rainfall intensities. The next significant peaks can be found in the 215–236th minutes, as it was mentioned. There were some sections where the length of the null value series was longer than the interval, here the correction factor was not allowed to be calculated (division by zero value of the 5-minute average intensity).

On the base of the experiences, the question marked intervals were not taken into consideration for further calculations. In the next part, the 10-, 20-, 30- and 60-minute $CF_5$ data will be investigated (*Fig. 2*).

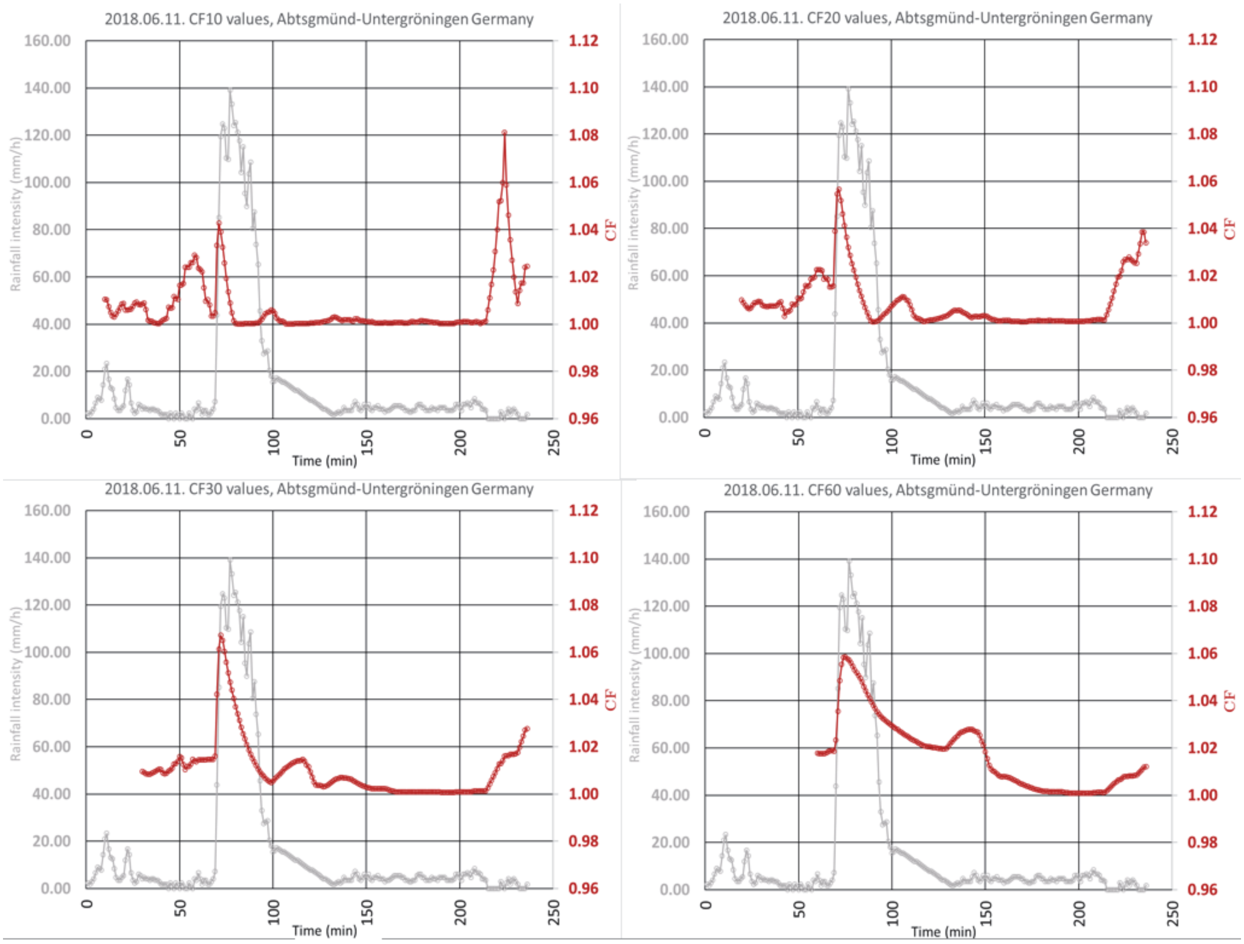


*Fig. 2*. The main characteristics of $CF_{10}$ (top left), $CF_{20}$ (top right), $CF_{30}$ (bottom left), and $CF_{60}$ (bottom right) values.

As the resulted values show, the magnitude of the correction is greater in longer sampling period cases. This is a logical consequence of the fact that as the sampling period is longer, the high values start to appear as peaks in the plain of the low intensity values, and their effect is longer, since the intervals are longer too. The highest demand of correction can be observed in the plot of the 60-minute data.

The minimum values of $CF_t$ are 1.00. The highest possible correction is influenced by the $b$ exponent. As the value of the exponent $b$ is less than 1.15 for the generally used TBG devices (*Vuerich et al.,* 2009), the maximums of the $CF_t$ values can be calculated. The maximum values are presented in *Table 1*. The magnitude of the possible maximum correction can be significant, but this is a theoretical value, as it was pointed out a little bit earlier, and it can occur in realistic cases rarely.

*Table 1*. Maximum values of $CF_t$ for various *b* values and some practical sampling periods

| b exponent | Sampling period | | | | |
|---|---|---|---|---|---|
| | 5 min | 10 min | 20 min | 30 min | 60 min |
| $b$=1.00 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $b$=1.05 | 1.084 | 1.122 | 1.162 | 1.185 | 1.227 |
| $b$=1.10 | 1.175 | 1.259 | 1.349 | 1.405 | 1.506 |
| $b$=1.15 | 1.273 | 1.413 | 1.567 | 1.666 | 1.848 |

Let us take a look at the averages of the $CF_t$ values of the sample time series. *Table 2* shows the average values of the given sampling periods, in case of $b = 1.042$. For the data of the 5-minute sampling period, the range of high fluctuating $CF_t$ values were excluded, where the measurement showed several consecutive nulls of rainfall intensities, so the maximum value is lower than in *Fig. 1*. As the data show, the average of the $CF_t$ values are relatively low, mainly in the shorter sampling periods. The correction can have significance in the 10-minute sampling periods and over, and mainly if the exponent is greater than 1.10, but for higher sampling periods, the correction can be verified even if the exponent has lower value.

*Table 2*. Average $CF_t$ values of the sample rainfall with a supposed *b*=1.042 exponent

| $b$ exponent | Sampling period | | | | |
|---|---|---|---|---|---|
| | 5 min | 10 min | 20 min | 30 min | 60 min |
| $b$=1.00 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $b$=1.05 | 1.002 | 1.009 | 1.010 | 1.012 | 1.021 |
| $b$=1.10 | 1.004 | 1.018 | 1.022 | 1.026 | 1.044 |
| $b$=1.15 | 1.007 | 1.028 | 1.034 | 1.040 | 1.069 |

The basic question is that in a $t$ sampling period with unknown $c_j$ weights, what is the value of the $CF_t$. As there are no data for the weights in a realistic case, an estimation is needed, and if it is possible, a generalized value should be used for the corrections in a certain geographic and/or climatic region.

For the determination of a generalized (or generalizable) value for the $CF_t$, the steadiness of its value is important, both in space and time. For this surmise, some initial hypothesis must be done. The first is that $CF_t$ values can 294haracterize a greater geographic region. This assumption is based on a likely steadiness of the $c_j$ weights, which can be similar in a wider region, independently from the kinds of rainfall, since even and quite varying weight distributions occurs in intensive and less intensive rainfalls, as well. The other hypothesis is that the weight characteristics of the rainfall were steady in time, so the distribution of weights were similar in the past, statistically. This surmise can be right, since there were similar types of rainfalls in the past, and as it was shown, the weights are sensible to the high ratio of the consecutive rainfall intensities, independently from their absolute value. Of course, these statements must be verified with the analysis of a great number of rainfalls in more geographic regions.

## *4. Conclusion*

In the paper, a method was presented as a simple tool for the correction of systematic biases of earlier measured long sampling period rainfall data, recorded by some known type of TBG rainfall gauges, where the sampling period was in the magnitude of 5, 10, 20, 30, and 60 minutes. The procedure was based on mathematical consideration, and the lack of detailed data was managed with the introduction of the generalized correction factor, the $GCF_t$, which can be calculated from short sampling period rainfall data (one minute), on the base of the supposed steadiness of the weight characteristics of consecutive one-minute rainfall intensities. It can be a good base of the correction presumably for wider geographical regions and longer time domain of measurements, since the method is based on the temporal distribution of weights of rainfall intensities in a unique sampling interval, without using the actual rainfall intensity values. The correction has significance in the 10-minute sampling period data, if the exponent of the correction equation of the TBG gauge is greater than 1.10, and for longer sampling periods, even for values greater than 1.05. The proposed method can help to clear the historical databases to make them a better reference for the investigation of IDF curves, and to make them a better reference for the analysis of the climate change relating to the rainfall intensities, and other parameters.

## *References*

*Adami, A.* and *Da Deppo, L.*, 1985: On the systematic errors of tipping bucket recording rain gauges. Proc. Int. Workshop on the Correction of Precipitation Measurements, Zurich. 1985. 1–3.

*Duchon, C.E.* and *Biddle C.J.*, 2010: Undercatch of tipping-bucket gauges in high rain rate events. *Adv. Geosci. 25*, 11–15. https://doi.org/10.5194/adgeo-25-11-2010

*Frankhauser, R.,* 1997: Measurement properties of tipping bucket raingauges and their influence on urban runoff simulation. *Wat. Sci. Tech. 36,* 7–12. https://doi.org/10.2166/wst.1997.0636

*Habib, E., Krajewsky, W.K.,* and *Kruger, A.,* 2001: Sampling errors of tipping bucket rain gauge measurements. *J. Hydraul. Engin.6,* 159–166. https://doi.org/10.1061/(ASCE)1084-0699(2001)6:2(159)

*Habib, E., Lee, G.,* and *Ciah, G.J.*, 2013: Ground-based direct measurement. In Rainfall: The State of Science, Washington: American Geographical Union, 61–78. https://doi.org/10.1029/2010GM000953

*Kurytka, J.C.* 1953. Precipitation measurements study. Urbana, Illinois, USA: State of Illinois, State Water Survey Division, Dpt of Education and Registration.

*Luyckx, G.* and *Berlamont, P.,* 2001: Simplified method to correct rainfall measurements from tipping bucket rain gauges. World Water Resources & Environmental Resources Congress, Orlando, 20-24 May 2001. Reston Va. USA: ASCE Publications. https://doi.org/10.1061/40583(275)72

*Marsalek, J.*, 1981: Calibration of tipping-bucket raingage*. J. Hydrol. 53,* 343–354. https://doi.org/10.1016/0022-1694(81)90010-X

*Mulvany, T.,* 1851: On the use of self registering rain and flood gauges in making observations of the relation of rainfall and flood discharges in given catchment. *Trans.Inst.Cio,Engrs.Ire, 4*, 18–33.

*Niemczynowicz, J.,* 1986: The dynamic calibration of tipping bucket raingauges*. Nordic Hydrol. 17*, 203–214. https://doi.org/10.2166/nh.1986.0013

*Vuerich, E., Monesi ,C., Lanza, L.G., Stagi, L.,* and *Lanzinger, E.,* 2009: WMO field intercomparison of rainfall intensity gauges. World Meteorological Organisation. Geneve.