

Application of Correction Procedures for Some Systematic Measurement Errors to Rainfall Intensity Data of a Rain Gauge in Budapest

Tibor Rácz^{1*}

¹ Doctoral School of Environmental Sciences, Hungarian University of Agriculture and Life Sciences, Páter Károly utca 1, H-2100 Gödöllő, Hungary

* Corresponding author, e-mail: raczt167@gmail.com

Received: 22 December 2020, Accepted: 10 May 2021, Published online: 27 May 2021

Abstract

The rainfall intensity measurement has a 150 years long history. In the first period of data recordings, the siphoned recording precipitation gauge (pluviographs), or siphoned rainfall writers (SRW), later, the tipping bucket gauges (TBG) were widely used. The systematic errors of these instruments resulted in lower intensity values for long periods. These errors were compensated sporadically. Most of the inaccurate data can be found in the high rainfall intensity range. Some of these data can be found in extracted, aggregated versions only, and the original measurement data is no longer available. These kinds of inherited systematic errors can be corrected. The fixing of siphoning error of SRWs and the supplementary correction of long sampling period data of TBG devices can be a suitable method for the elimination of these issues. In this paper, the application of these two methods is shown in a case study to point out the magnitude and effect of these errors on the IDF curves. The case study on the use of the before-mentioned correction procedures is performed on the rainfall data of the Budapest-Belterület (Budapest City) rainfall station, using data series spanning 105 years. These corrections show that the earlier IDF curves can show 5–10% lower intensities, mainly in the short and low return frequency rainfalls. The result of these kinds of corrections can be significant for the climate change investigations or in the re-evaluation of the elder IDF curves.

Keywords

rainfall intensity, accuracy, siphoning gauge, Tipping bucket gauge, data correction, older rainfall data

1 Introduction

The importance of the rainfall intensity data increased after the invention of the rational method in 1851, by Mulvany [1]. His result became the principle of the planning of rainfall drainage systems. In Hungary, in the second part of the XIXth century the urban rainwater drainage issue had great importance, parallelly with the development of sewer drainage systems. One of the key variables of the rational method is the rainfall intensity, and as the demand of rainfall data had grown, the meteorological data collecting developed as well. In the Hungarian Kingdom there had been rainfall data collection, but only sporadically since the middle of the 1700s. The foundation of the antecedent of the Hungarian Meteorological Service (Royal Hungarian Central Institute of Meteorology and Earth Magnetism) in 1870 can be considered as the beginnings of the systematic measurement in Hungary [2]. Recently, the

rainfall intensity data of the last 20 years were published for 101 automatic operation rain gauge stations to help the work of engineers in the changing climate [3].

Before the 1990s, the data were collected mainly using siphoned pluviographs (SRW), and before the second world war, additionally weight measurement-based pluviographs were in use (Andorkó-Bogdánfy rain gauge). As a part of the automatization of the rainfall measurement, in the 1990s the SRW devices were changed into tipping bucket rain gauges (TBG).

The rainfall intensity measurement methods have their specific errors. A part of them is similar to the measurement issues of the traditional rain gauge's errors, such as the inaccuracy caused by wind, the wetting and the splashing. The correction of these errors is solved basically for multi-day rainfall data, but for sub-daily data, there are no correction

procedures. Over these issues, there are also systematic errors, depending on the method of measurement technique of a given technique; this article focuses on these errors.

The systematic error of TBG has been investigated since the 1940s, more often since the 1980s. The main result of the research was the ascertainment of the negative error of measurement, increasing with the rainfall intensity [4–8]. The non-linearity of the error with rainfall intensity was observed earlier, and so polynomial and power function was proposed for the correction [9]. In this study, the power function is used on the results of Vuerich et al. [10] and Lanza et al. [11]. In their work, the correction was developed for one-minute sampling period data. A significant part of the older rainfall data of TBG devices is given in 10-minute period, as the sum of the one-minute measures of the 10-minute interval [10], or in case of elder data bases there are not one-minute data at all, for the further processing. In the case of the TBG devices, the source and measure of the error was investigated, the correction of the short sampling period data was solved. The TBG data in the first period were collected in long sampling periods because there were not adequate data storage or data transmission solutions in one minute-base. The correction of the long sampling period data requires a supplementary correction, depending on the accuracy of the gauge.

The systematic error of the SRW devices was studied as well, for the correction on the registration tapes. First of all Luyckx and Berlamont [12] developed a method for the correction of the systematic error of SRW devices, if the registration chart is available. In the archives, there are mainly processed SRW data from the early times. The SRW registration ribbon were analyzed several decades ago, and only the maximum intensity values were collected to a chart by specific time windows. The registration chart was put back to the archive, and some of them cannot be find again anymore. The most extended sampling period was 10 or 15 minutes.

For the correction of the systematic errors of the SRW and TBG data, two methods were elaborated [13, 14].

This article shows the effect of the correction of the mentioned systematic errors on the rainfall intensity data, and on the IDF data, as well.

2 Data, a short introduction of rainfall intensity measurement in Budapest

The data for the present research were collected at Budapest Belterület (N47.5109 E19.0280) rainfall measurement station (henceforth referred as Station) between 1915 and

2018 (Fig. 1, [15]). The Hungarian Meteorological Service (HMS) has given free run of the data partially in scanned handwritten processing sheets (SRW data of a Hellmann-Fuess device) and in a data file (TBG data of a Lambrecht 15188 device).

The SRW data seemed worthy to be divided into three parts since there was a longer break in the data between 1929 and 1941, and in 1967 there was a change in the measurement when the rainfall recording was extended to the whole year, meanwhile in the earlier periods it was limited to the summer months. It was probable that in this occasion the device was changed, as well. Hence, the SRW data were divided into the D1, D2A, D2B subsets, respectively. The processed SRW data of these subsets comprise the highest rainfall depth of the 5-, 10-, 20-, 30-, 60-minutes time intervals for several intensive rainfalls in a year, more detailed data were not available from these periods.

The set of the TBG data were signed with D3. These data were collected in every 10 minutes. Comparing to the continuous data of the SRW records, this kind of sampling could have caused a probable loss of data. This error is the consequence of the sampling theorem since the change of the rainfall intensity is quite fast than this length of period. The most intensive 10 minutes period can be caught in ten different ways with the same probability, and only one would give the possible maximum configuration (if the peak values occur in a shorter time), so the maximum can be pointed out only with 0.1 probability with the 10 minutes sampling. This processing error can cause a 5–10 % data loss on average [16]. In this work this error was not taken into consideration. The composition of the available data is shown in Fig. 2. The subsets of data opened the possibility of an analysis of the identicalness of the whole dataset.



Fig. 1 Gauge garden of the Hungarian Meteorological Service in the 1950s (source: www.met.hu)

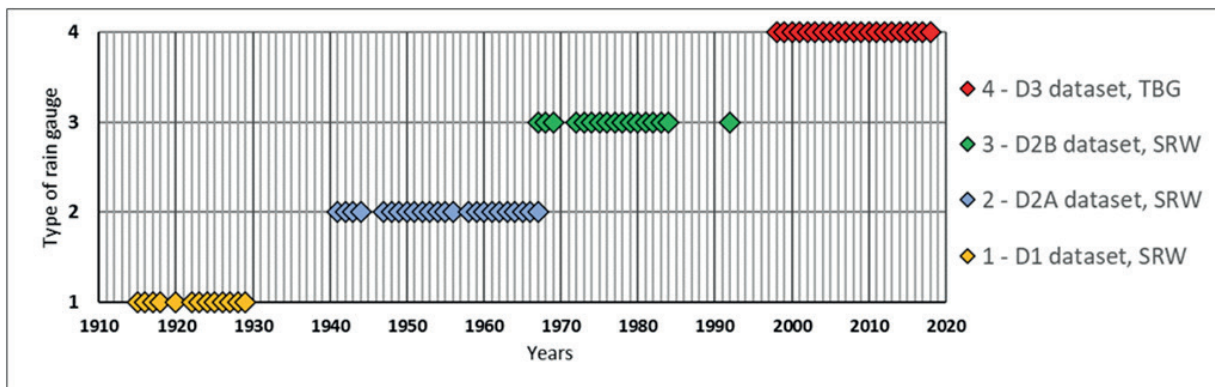


Fig. 2 Available rainfall dataset with its subsets (Budapest Belterület station, 1915–2018 period)

The site of the meteorological Station is at the central building of the Hungarian Meteorological Service. This site is a plot in an urbanized district, where the built-in rate developed during 20th century. In the 1930s the neighborhood plots were built-in with 4–5 storied buildings. In the 1980s, the plot of the measurement garden was sold, so the observation must have continued in new places. The SRW devices were left on the terrain, until the end of their service. The TBG device was placed on the terrace of the building where the traditional rain gauges were positioned earlier. The TBG device was in operation undisturbed during the whole period of the measurement, despite of the reconstruction works performed on the pavement of the terrace, when other devices were moved to other places temporarily.

3 Methods

In the first part of data processing, the following operations and investigations were done.

First of all, the examination of the data quality was necessary to find the seemingly inappropriate data. After that the basic cleaning or correction was needed.

The second step is the correction of the errors originating from the applied measurement methods, and the third was the checking of the homogeneity of data. After these tests, the data were modelled using an adequately selected probability distribution. The yearly exceeding probability of the intensity for the unique durations was estimated on the basis of the fitted probability distributions. On the estimated probabilities of the rainfall intensities, probability-based IDF curves, and their power function approaches were generated for raw and corrected data, and a comparison was performed to investigate the effect of the correction.

For the evaluation of the data, there are two widely used ways. The first is the investigation based on the annual

maxima (AM) approach, so the set of yearly maximum values of rainfall intensities of different durations can be inspected. The second way is the Peak Over Threshold (POT) analysis where the object of examination are the data which succeed an adequately chosen threshold value. In this investigation the AM method is applied since there are relevant differences between the collection and processing of the original data subsets.

The SRW and TBG gauges result in different data structures. Initially, the SRW data were in the form of analogous registration charts which were processed on paper sheets several decades ago, showing the rainfall depths fallen in the most intensive 5-10-20-30-60 minutes for the most significant showers of a unique year. In this work, the maximum data were selected for the further processing from the sheets. In the case of the SRW data, the highest sums of rain depths were transformed into intensities. The data measured by TBG device have a 10 minutes sampling period. These data were processed to gain the highest values of the measured 10-, 20-, 30-, 60-minutes rainfall depths similarly to the SRW data, and the transformation of rain depths to intensity was similarly done. There is a significant difference between the data of the SRW and TBG databases. The SRW data can catch the highest intensity intervals. The TBG device is constrained to the sampling period, so there is the probable underestimation which comes from the earlier mentioned processing error [16]. This difference can result in a certain underestimation in the TBG data.

In the case of the SRW devices, there is a source of the error caused by the siphoning process, resulting in a systematic undercatch. This error can be corrected by the Eq. (1) [13], which is a developed correction method of Luyckx and Berlamont [12] for the correction of stand-alone interval's intensities.

$$i_{corr} = \frac{\left(t + \text{int} \left(\frac{h_0 + t \cdot i_t}{h_s} \right) \left(\frac{t_{s,0}}{1 - \frac{A_f \cdot i_t}{q_s}} \right) \right)}{t} \cdot i_t, \quad (1)$$

where i_{corr} is the corrected rainfall intensity for a t long time interval, the t is the length of the time interval, the i_t is the raw (not corrected) rainfall intensity for the t time window, h_s is the measurable rain depth between two consecutive siphoning processes (width of the registration chart), h_0 the initial water level in the device (the value of h_0 is $h_s/2$, as the expected value of h_0 what follows a uniform distribution), $t_{s,0}$ is the duration of a siphoning process, A_f is the catching area of the SRW, and the q_s is the siphoning rate of the device, the $\text{int}()$ function is the integer part of the expression in the parentheses. This function counts (estimates) the number of siphonages during the t time. The factor of the i_t expresses the ratio of an extended time of the measurement interval over the real time. The numerator comprises the sum of t and the estimated time of siphoning during the given interval. The value of this factor is at least one, if there were siphonage during the t period, an increment is calculated with the i_t intensity, using the hydraulic characteristics of the device.

The TBG devices have a systematic undercatch in the range of the high intensities and some cases a slight overestimation in the low intensities. The correction of the TBG measured data was performed according to the method published by Vuerich, et al. [10] and Lanza et al. [11]. On the basis of field experiments of Lanza and his colleagues, a correction equation was made which describes the relationship between the reference values of pit gauges and the measured values on several devices. For the Lambrecht 15188 gauge, the following equation can be gained by the transformation of the presented regression:

$$i_{corr} = a \cdot i_t^b = 0.82 \cdot i_t^{1.042}, \quad (2)$$

where i_{corr} is the corrected rainfall intensity, and i_t is the raw (not corrected) rainfall intensity. Since the Eq. (2) was determined according to one-minute based calibration, a correction is needed to get the ten-minutes fixed data. The reason of the requirement of a further correction is the non-linearity of the systematic error. The equality of rainfall amount measured in ten consecutive one-minute measurement and one ten-minute sampling is evident for a given time. However, if a non-linear correction is performed on the one-minute measurements and on the

ten-minute measurement, the corrected amounts will be different. For the correction, a method was elaborated expressing the ratio of the fixed rainfall amounts calculated from the one-minute and ten-minute corrections with weight numbers to express the inequality of rainfall amounts of the one-minute data in the 10 minutes time interval. On the basis of this deduction [14], the following correction formula was used:

$$i_{corr,10} = \frac{\sum_{i=1}^{10} (c_i^b)}{10} \cdot 0.82 \cdot i_t^{1.042} = CF_{10} \cdot 0.82 \cdot i_t^{1.042} \quad (3)$$

where the $i_{corr,10}$ is the fixed 10 minutes rainfall intensity, the c_i values are weight numbers, which express a relation of the one-minute intensity to the 10-minute average intensity, and CF_{10} is the correction factor of the one-minute rainfall intensity for the upgrade to 10 minutes sampling period.

The CF_{10} factor depends on the distribution of the one-minute rainfall intensities. Since the one-minute data is not known in our case, the CF_{10} value must be estimated. The c_i weights are bounded values in the $(1, n^{b-1})$ closed interval, so they depend only on the pattern of the rainfall intensities of a certain time interval, independently of the magnitude of the intensities. It means that these weights are more robust quantities than the rainfall intensities, and they can be characteristic for the rainfall, in general; presumably in wide temporal range and geographical region [14]. The geographic distribution of the CF_{10} values are not investigated; the simplest way of their estimation is the use of some local rainfall data.

To determine the correction factor, an analysis on data from an extreme intensity rainfall event was performed to get the average CF_t values within class t (unpublished result). According to this result, CF_t values increase with increasing t . Their values are also determined by the exponent b as shown in Eq. (3). For the Lambrecht 15188 instrument, the CF_t value for the 10-minute intervals was 1.007. The TBG correction was therefore made using this value, which in this case could have very little significant effect. However, there are more widely used instruments in practice where b is around 1.1 when CF_t is 1.015, which may result in a meaningful improvement in the data.

Since at least two kinds of measurement techniques were applied with at least three devices, a homogeneity test was necessary to check the effect of the mentioned changes, or the impact of the moving of the observation site from the courtyard of the central building of the Meteorological Service onto its roof terrace. For the

homogeneity test, the XLSTAT package was used, under the Excel [17] which operates four kinds of tests: Pettitt, SNHT, Buishand and von Neumann tests. The Buishand and SNHT were not suitable in the case of the presented data, since these tests assume the normality of the sample, meanwhile the Pettitt and von Neumann tests can be used for any kind of distributions [17]. The yearly maxima show an asymmetric distribution, therefore the Pettitt and von Neumann tests were used in this work.

The Pettitt test is a non-parametric test which determines and locates the date of the supposed breakpoint; its null hypothesis is that the data follows one or more distributions with the same location parameter. The alternative hypothesis is that there is a time where the location parameter changes [17, 18], in the investigated significance level. Since there can be breakpoints in every data series in different significance levels, the probable breakpoint is pointed out, even if the null-hypothesis has not to be rejected. The Pettitt test does not detect the change in distribution if there is no change of location parameter [17].

The von Neumann test is a non-parametric test to examine the null hypothesis whether the data are independent, identically, and randomly distributed; nevertheless, the test does not provide information about the possible breakpoint [18]. The von Neumann test detects the change without information about its location [19]. The homogeneity test can show that existence of an identical distribution for the data series of 5-, 10-, 20-, 30-, 60-minutes maximum rainfall intensities. As a double-check, the identicalness of the D1, D2A, D2B, D3 parts of the data series were also investigated. For this statistical testing, the non-parametric Kruskal Wallis H test was applied, which is a tool for the comparison of the identicalness of the distribution of more samples in one procedure [17].

After the verification of the identicalness of the samples, probability distributions were selected to model the dataset. To define the distribution of the maxima of the rainfall intensities, the Extreme Value Theory (EVT) was used since the samples comprise the collection of the yearly maximum values of rainfall intensities, which were proven as identically distributed random variables. The Fisher-Tippit-Gnedenko Theorem results in the applicability of one of the Gumbel, Weibull or Fréchet distributions to model the rainfall intensities [20]. These extreme value distributions are the specially parametrized cases of the generalized extreme value distribution (GEV), so the resulted empirical cumulative distributions were modelled with GEV.

The fitting was performed with the EasyFit statistical software. The EasyFit applies four methods for the distribution fitting, the method of moments (MOM), the maximum likelihood estimates (MLE), the least squares estimates (LSE) and the method of L-moments. EasyFit uses the least computationally intensive methods for the given distributions. If there is a possibility of estimation of parameters of moment estimation, the method of moments is used to avoid the iterative numerical methods. For most of distributions the iterative parameter estimation algorithm is used, first of all the MLE method involving the maximization of the log-likelihood function.

For the GEV distribution, EasyFit implements the numerical method for multi-dimensional function minimization. The algorithm terminates when the stopping criteria is satisfied (the specified accuracy of the estimation is reached, or the number of iterations reaches the specified maximum).

The goodness of fitting was checked with Kolmogorov Smirnov (KS), Andersson-Darling (AD) and Chi-squared (CH) tests.

Using the probability distribution, the IDF curves can be obtained, and the difference between the raw and corrected data can be studied.

4 Results and discussion

4.1 Data cleaning

The D1, D2A, D2B and D3 datasets were checked on the basis of the scaling characteristic of the rainfall intensity. This principle results in the maximum of a smaller time scale which cannot be lower than the maximum of the larger scale in one rainfall event. There were four questionable data in the database, and these were substituted by estimated values, on the basis of the power function

$$i = a \cdot t^b, \quad (4)$$

where i is the rainfall intensity, t is the time scale (length of the time window), and a , b are parameters, determined on the basis of the neighboring intensity data.

4.2 Correction of the systematic errors

The correction of systematic undercatch of SRW and TBG devices results in a slight increase of the intensity data. The higher is the intensity, the correction is more significant.

The highest rate of correction occurs in the 5 min data; the highest correction of unique data in the SRW was 13 %, meanwhile the mean of the corrections was 5 %.

The discrete characteristics of the SRW correction can be seen in the Fig. 3, where the lower range of intensities does not show modifications of raw data. This is the consequence of the structure of SRW gauges, since in the case of low intensities the emptying does not occur. There were no TBG data for this duration, since the sampling period of the TBG device was 10 minutes.

In Figs. 4 and 5 the effect of the correction of the longer durations can be seen.

In the 10 minutes duration data, the highest correction was 10 % and the average of the corrections was 3 %. As the duration increases, the magnitude of the correction decreases. The maxima of the corrections are in a similar magnitude in both of correction procedures and devices. The maximum corrections of TBG and SRW data are shown in Table 1.

The effect of the correction can be studied, of course, in the empirical cumulative distribution, as well. Fig. 6 shows the empirical distributions for the raw and corrected data of every investigated durations. It can be seen in the five-minute data, that the siphon starts to operate only above a certain threshold, and so no correction occurs for lower values. However, according to the approximation

used in Eq. (1), the threshold is observed at an intensity of 60 mm/h due to the initial water level chosen at 5 mm in the middle of the measuring range of the device. This threshold can be found at increasingly lower intensities for longer t intervals. If the initial water level were not taken into account, the threshold for the 5-minute correction would be higher, with corrections taking place at a higher threshold. In such cases many corrections would be missed. A statistically better correction can be made with an initial value of 5 mm.

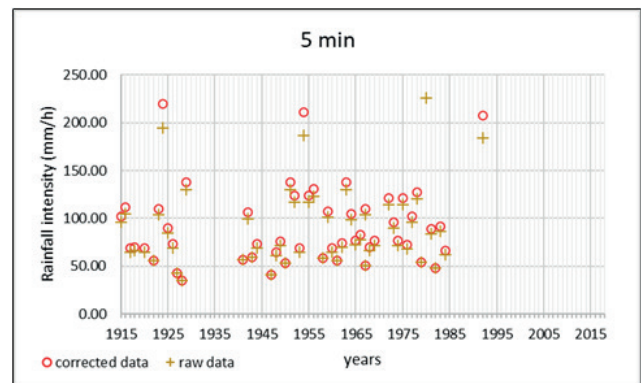


Fig. 3 Data correction of 5 min duration rainfall intensity data

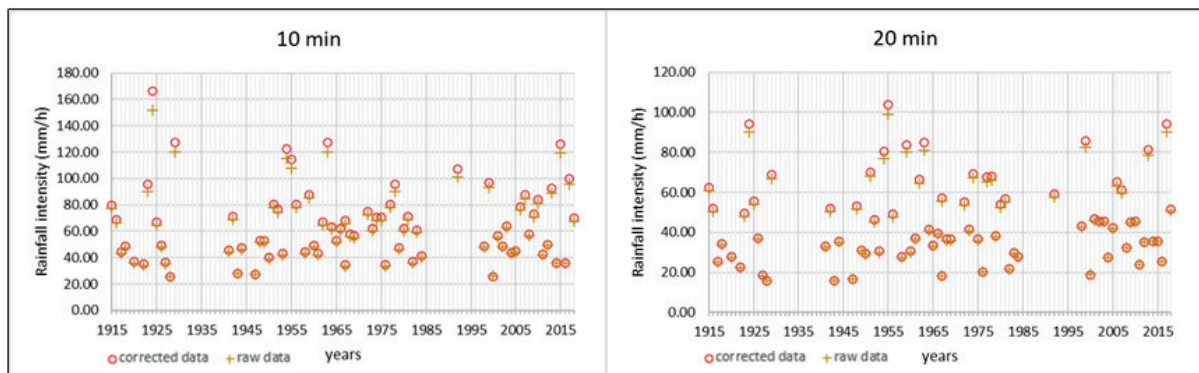


Fig. 4 Data correction results of 10 and 20 minutes duration rainfall intensities both of SRW and TBG measures

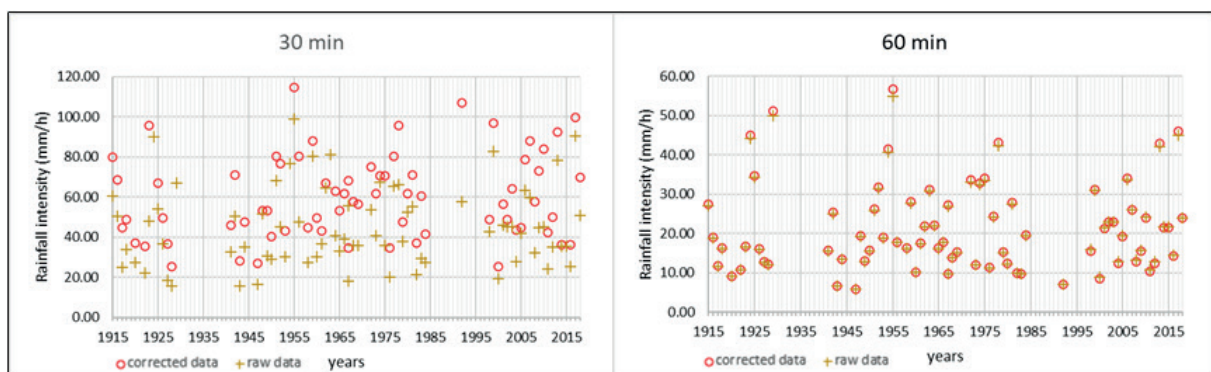


Fig. 5 Data correction results of 30- and 60-minutes duration rainfall intensities both of SRW and TBG measures

Table 1 Maximum corrections of TBG and SRW data

<i>t</i> durations	10 min	20 min	30 min	60 min
TBG data (%)	5.4	4.4	4.0	2.3
SRW data (%)	9,7	4.7	4.1	3.1

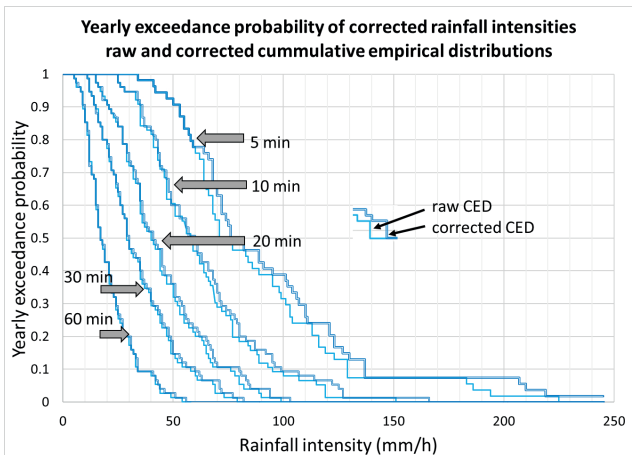


Fig. 6 Cumulative empirical distributions of the raw and corrected data for different sampling periods

The greater differences between the corrected and the raw data can be observed at the 5-, 10-, 20-minutes intensities. For the 30-, 60-minutes time window, the rate of the correction is not significant, and the raw and corrected data are close to each other. Overall, it can be stated that the underestimations caused by systematic errors could have been corrected.

4.3 Verification of homogeneity

In the homogeneity test, the 5-, 10-, 20-, 30-, 60 minutes complete data series were examined. It means that this investigation did not take care of the different measurement conditions, such as the different techniques, or site, or the sampling circumstances. The test would show the casual traces of inhomogeneities, as the consequence of the changing measurement environment and condition. The operation was performed with XLSTAT software, using the Pettitt and von Neumann tests [9]. The missing values were neglected. The result of the homogeneity test showed that in the data series there were no breakpoints which would have caused inhomogeneity, on the basis at 5 % significance level. The resulted p-values of the homogeneity tests are shown in Table 2.

As the p-values are higher than 0.05, the data series of the different durations are homogeneous as both tests showed. The *t* values in the Pettitt test sign breakpoints in 1951 which has no effect on the homogeneity in this confidence level, since the null-hypothesis was not rejected.

This breakpoint can be the result of a period when the daily maximum precipitation was relatively low, lower than 60 mm.

4.4 Verification of the identicalness of the sub-samples

Although, the homogeneity of the data series was verified by the Pettitt and von Neumann tests, as a double-check, the data subsets of each time windows were investigated. As there are more than two samples, the Kruskal Wallis H test was a proper tool for the investigation of identicalness [17]. The test is a built-in module of the XLSTAT package. The results of the test are presented in Table 3.

The result of the Kruskal Wallis H test (Fig. 7) proved at 5 % significance level that the D1, D2A, D2B, and D3 data subsets are identical, and they can be modelled by the same distribution. The primary statistical parameters of the data subsets are presented in Fig. 6. The boxplot shows the spreading of the sample; the red crosses sign the means of the samples. Although, these samples are not really long, and there are gaps between the datasets, the uniformity of means stands out. None of the classes of rainfall intensity maxima show significant increasing or decreasing. This result can sign that there is no significant trend in the yearly maximum rainfall intensity data in this approximately 100-year time span data set.

The lack of some significant trend can be the result of the geographic situation of the Station; since Budapest is relatively far from the seas, the major part of the moisture transported by wet air from the Mediterranean region

Table 2 Evaluation of homogeneity tests

Time frame in minutes	5	10	20	30	60
Pettitt test					
K	229	300	297	284	248
T	1951	1951	1951	1951	1951
p-value (two tailed)	0.416	0.910	0.938	0.946	0.629
Alpha	0.05	0.05	0.05	0.05	0.05
von Neumann test					
N	1,993	2.132	1.993	2.093	1.918
p-value (two tailed)	0.574	0.714	0.391	0.512	0.274
alpha	0.05	0.05	0.05	0.05	0.05

Table 3 Evaluation of Kruskal-Wallis H test

Time frame in minutes	5	10	20	30	60
Kruskal-Wallis H test parameters					
K	0.261	0.173	0.453	0.402	1.496
p-value (one-tailed)	0.881	0.982	0.937	0.941	0.678
Alpha	0.050	0.050	0.050	0.050	0.050

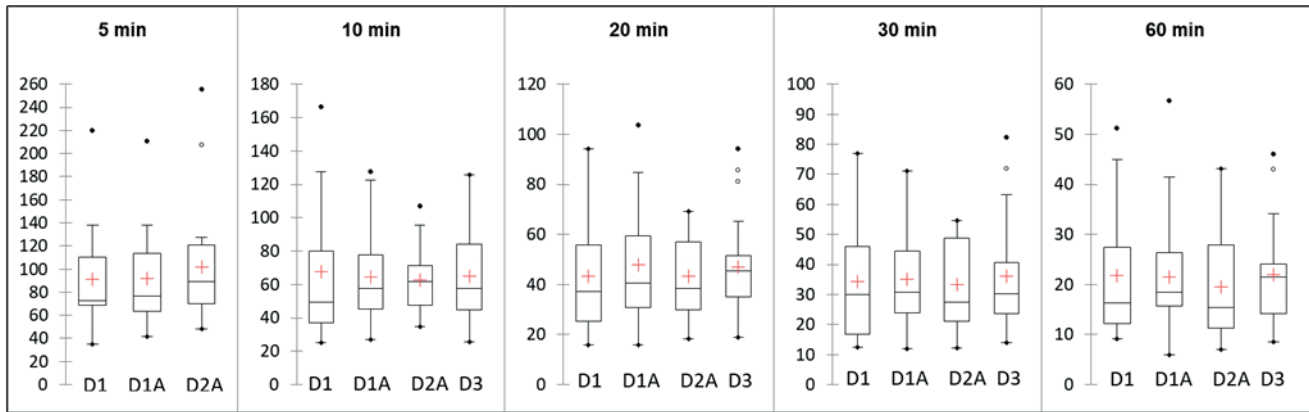


Fig. 7 Statistical parameters of Kruskal Wallis H test

falls out during the thousand kilometers path, and so the rainfalls cannot cause significant rainfalls, despite of the increased evaporation, as a result of the augmented seasonal average and peak temperature in the past decades in this geographical region. This hypothesis can be the object of further studies.

There is no contradiction to the result of other studies which found the decreasing number of rainy days especially in summer (this is the most wet period in Hungary) and increasing daily average rainfall intensities [21], because these investigations do not tell anything about the peak intensities of unique rainfalls.

The frequency of rainy days is not related necessarily to the maximum intensity of precipitation.

4.5 Probability distribution fitting

The results of the fitting of GEV distribution on the complete rainfall intensity dataset can be seen in Fig. 8.

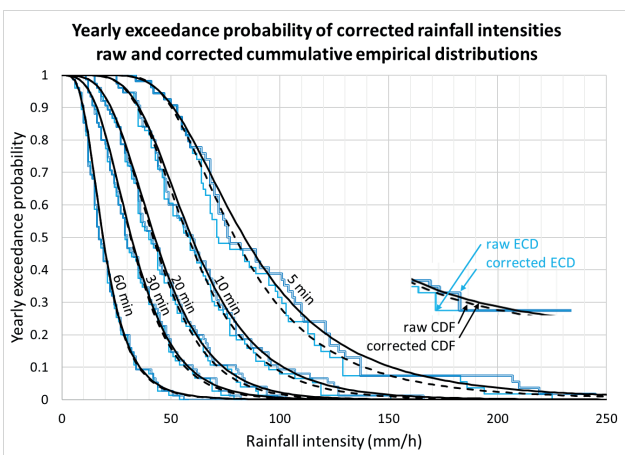


Fig. 8 Yearly exceedance probability distribution models of the rainfall intensities on the complete raw and corrected dataset (D1+D2+D3) of the yearly maximum rainfall intensities with the empirical cumulative distribution functions (ECD)

The goodness of fitting was checked with the Kolmogorov Smirnov, Andersson Darling and Chi-Squared tests at 5 % probability level.

The appearance of the fitted distributions of 5 minutes raw and corrected curves shows overlapping in the low and moderate range of rainfall intensities. This is the consequence of the lack of TBG measurements in this duration category. As it was mentioned, the SRW corrections occurred only in a higher range of the intensities, and the ratio of the corrections decreased for the longer durations.

The most characteristic difference between the raw and corrected data can be found in the 10-minutes intensities. This is the result of the relatively high intensities since the correction is generally higher than in the cases of the longer duration data. In the case of the appearing TBG data, which has a continuous correction formula, there were corrections for every data. This continuous correction balanced the SRW data where the corrections must have been performed only above a threshold value, similarly to the case of the 5 minutes intensities.

In the case of the 20-, 30-, and 60-minutes data, lower difference can be seen between the exceedance probabilities of the raw and fixed data. In the lower range, the correction of the TBG data has a downward tendency, and its ratio to the raw intensity is lower. The occurrence of the SRW corrections is less probable in the low range of the intensities.

Generally, the correction has significant effect in the range of the high intensities and rare rainfall events.

4.6 IDF curves

The IDF data were calculated on the basis of the model distribution. The raw data are shown in the Table 4, the corrected data in the Table 5, and the Table 6 shows the ratio of the corrected and raw values.

Table 4 IDF data - Raw data (mm/h)

Average Return Frequency (years)	5 min duration	10 min duration	20 min duration	30 min duration	60 min duration
1	35.6	21.5	10.8	6.9	5.1
2	78.8	57.5	41.5	31.3	18.6
4	104.1	75.3	55.5	42.9	25.9
10	138.2	96.5	71.4	56.4	35.1
20	166.4	112.2	82.7	66.2	42.3
50	208.6	133.4	97.3	79.1	52.4
100	244.8	149.8	108.0	88.9	60.7

Table 5 IDF data - Corrected data (mm/h)

Average Return Frequency (years)	5 min duration	10 min duration	20 min duration	30 min duration	60 min duration
1	34.8	21.9	10.7	7.6	5.0
2	83.1	59.1	42.2	31.5	18.7
4	111.8	78.1	57.0	43.4	26.1
10	150.6	101.2	73.9	57.7	35.6
20	183.0	118.7	86.2	68.4	43.1
50	231.7	142.7	102.0	82.7	53.7
100	274.0	161.7	114.0	94.0	62.5

Table 6 IDF data ratio of Corrected and Raw data (-)

Return Frequency (years)	5 min duration	10 min duration	20 min duration	30 min duration	60 min duration
1	0.98	1.02	0.99	1.11	0.99
2	1.05	1.03	1.02	1.01	1.00
4	1.07	1.04	1.03	1.01	1.01
10	1.09	1.05	1.03	1.02	1.01
20	1.10	1.06	1.04	1.03	1.02
50	1.11	1.07	1.05	1.05	1.03
100	1.12	1.08	1.06	1.06	1.03

The ratio of the increment generally increases with the return frequency, the only exception can be found in the case of the 60 minutes data. This irregularity can be explained by the less accurate fitting of the distribution at the tail.

The increment of the 5-, 10-, 20-, 30-minutes durations rainfall intensity values, in the 10-100 years frequency category is significant, its magnitude is in the 7–11 % interval. In the urban drainage tasks, for the simple traditional drainage systems, the 2–4 years return frequency is the generally expected planning target, the higher return periods have importance in the planning of large span collector drains. The increment of the intensities, as a consequence

of the correction of the systematic measurement errors can explain the overload of earlier laid rainfall drains for extreme storm events, in some cases.

The IDF data were plotted in the Fig. 9. The curves can be approximated with power function, as it is common in the practice, in most of the countries.

The power functions fitted to the intensities have the a and b parameters by the Eq. (4). These parameters were determined by EXCEL and are shown in the Table 7.

In the practice these power function curves are used, but there are significant differences between them and the IDF data derived from the probability distributions. The highest differences can be found in the 5- and 60-minutes intensities, and in the 1 % exceedance probability data, for almost all t intervals.

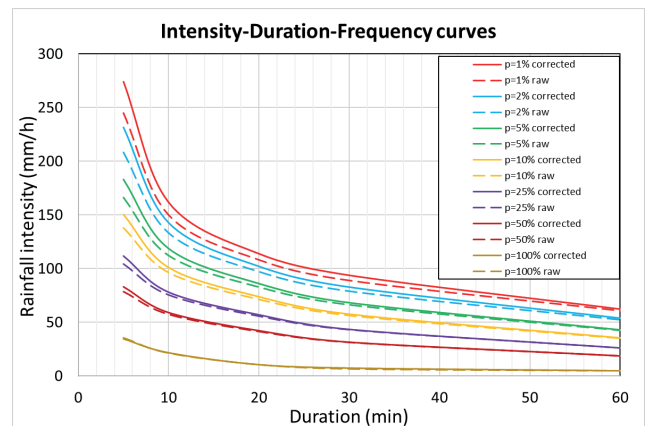


Fig. 9 Rainfall IDF data on the basis of the GEV probability distribution; the uncorrected intensity data are presented by the dashed lines, the continuous lines show the corrected rainfall intensity data

Table 7 Parameters and of IDF curves

Average Return Frequency	Exceedance Probability		a	b
1	99	raw	564.07	-0.55
2	50	raw	488.07	-0.54
4	25	raw	398.79	-0.54
10	10	raw	338.42	-0.54
20	5	raw	265.24	-0.55
50	2	raw	210.91	-0.57
100	1	raw	133.49	-0.83
1	99	corr.	659.01	-0.58
2	50	corr.	564.96	-0.57
4	25	corr.	455.50	-0.57
10	10	corr.	382.04	-0.57
20	5	corr.	293.62	-0.57
50	2	corr.	228.23	-0.59
100	1	corr.	129.62	-0.81

4.7 Uncertainties of the correction

In this paper, the characterization of the accuracy of SRW and TBG correction can only be qualitative, since there are no available temporally detailed data for a comparison of the investigated period. Despite of this issue, some estimation can be done to characterize the magnitude of the possible inaccuracy of the applied correction methods.

The accuracy of the SRW is the question of the difference between the actual intensity at siphonages and the mean intensity of the given interval. The shorter is the interval, the closer are the intensity values to the mean intensity at siphoning.

The accuracy of the TBG correction depends on the length of the t , which is n minutes long. In a realistic case, there are rainfall in every minute of the t interval, so the correction factor cannot reach its possible n^{b-1} maximum value. For long t and greater n , the correction factor is higher since the base of the power is greater. It means that the necessity of TBG correction is higher if the sampling interval is greater. Unique correction factor can be used for every t intervals, but it needs a longer analysis to determine its reliable value.

5 Conclusions

The present study has shown the application of the correction methods of the systematic errors of SRW and TBG devices for earlier processed data. Two methods were used to fix the systematic error in databases which comprise processed SWR data or long sampling period TBG data. The correction of the SRW data is limited for those datasets which contain only selected maximum data of time windows of unique rainfalls. The model applied for the calculation takes into consideration the supposed initial water level of the rainfall recorder with its expected value, and by this way, a good estimation can be done to the necessary corrections.

The correction of the TBG data is continuous, since the systematic error happens in every tipping, only its extent is higher when the rainfall intensity is higher. For those datasets which were registered with long sampling period, the correction formulas must be completed with a

supplementary factor. After performing the correction, the increase of the rainfall intensities can be experienced. In this data, since a good performance rain gauge was used, the supplementary correction was quite low. This correction is proposed mainly for those devices which show significant deviation in the measured data, in the higher range of rainfall intensities.

In this data, the experienced maximum correction was around 10 %, in the highest intensities of shorter intervals and rare occurrences.

Regarding the analyzed 105-year span dataset, the statistical tests showed homogeneity. It means that this dataset does not show significant changes. The reason of this result can be the higher robustness of the rainfall intensity against the rainfall quantity measurements. This claim demands more investigations and proves.

The correction of the SRW and TBG data strongly depends on the technical parameters of the gauges, such as the catching surface of the funnel, time of emptying, discharge of emptying at SRW instruments, and the a and b correction parameters for the TBG devices, as it can be seen in the Eqs. (1–2). Although, there remain some other relevant uncertainties in the rainfall data, such as the wind effect, but at least the systematic errors of earlier measured, processed data can be handled successfully, using the presented methods. To achieve this aim, several elderly used instrument's calibration data must be collected or re-measured.

Using the presented corrections, old databases can become better reference for the comparison of the effects of climate change on the rainfall intensities, as well.

Acknowledgement

This article could not have been done without the help of the Hungarian Meteorological Service, which provided the archival precipitation intensity data of the Budapest Belterület station and the digital data file of the automatic precipitation meter operating at the Station for the last two decades. I thank my friend and colleague Laura Varga for the digitalization of the data of the siphoned pluviograph of this Station.

References

- [1] Mulvany, T. J. "On the Use of Self-registering Rain and Flood Gauges in Making Observations of the Rainfall and Flood Discharges in a Given Catchment", Transactions of the Institution of Civil Engineers of Ireland, 4(2), pp. 18–33, 1851.
- [2] OMSZ "History of OMSZ", [online] Available at: https://www.met.hu/en/omsz/OMSZ_tortenete/ [Accessed: 22 October 2020]
- [3] OMSZ "Csapadékinzintitás" (Rainfall intensity), [online] Available at: <https://www.met.hu/eghajlat/csapadekintenzitas/> [Accessed: 04 March 2020]

- [4] Adami, A., Da Deppo, L. "On the systematic errors of tipping bucket recording rain gauges", presented at the Workshop on the Correction of Precipitation Measurements, Zurich, Switzerland, 1985.
- [5] Marsalek, J. "Calibration of tipping bucket raingauge", *Journal of Hydrology*, 53(3–4), pp. 343–354, 1981.
[https://doi.org/10.1016/0022-1694\(81\)90010-X](https://doi.org/10.1016/0022-1694(81)90010-X)
- [6] Niemczynowicz, J. "The Dynamic Calibration of Tipping-Bucket Raingauges", *Hydrology Research*, 17(3), pp. 203–214, 1986.
<https://doi.org/10.2166/nh.1986.0013>
- [7] Frankhauser, R. "Measurement properties of tipping bucket rain-gauges and their influence on urban runoff simulation", *Water Science and Technology*, 36, pp. 7–12, 1997.
[https://doi.org/10.1016/S0273-1223\(97\)00625-2](https://doi.org/10.1016/S0273-1223(97)00625-2)
- [8] Habib, E., Lee, G., Kim, D., Ciach, G. J. "Ground-based direct measurement", In: Testik, F. (ed.) *Rainfall: The State of Science*, American Geographical Union, Washington, DC, USA, 2013, pp. 61–78.
- [9] Vasvári, V. "Calibration of tipping bucket rain gauges in the Graz urban research area", *Atmospheric Research*, 77, pp. 18–28, 2005.
<https://doi.org/10.1016/j.atmosres.2004.12.012>
- [10] Vuerich, E., Monesi, C., Lanza, L. G., Stagi, L., Lanzinger, E. "WMO field intercomparison of rainfall intensity gauges", *World Meteorological Organization*, Geneva, Switzerland, 2009.
- [11] Lanza, L. G., Vuerich, E., Gnecco, I. "Analysis of highly accurate rain intensity measurements from a field test site", *Advances in Geosciences*, 25, pp. 37–44, 2010.
<https://doi.org/10.5194/adgeo-25-37-2010>
- [12] Luyckx, G., Berlamont, J. "Accuracy of siphoning rain gauges", presented at Ninth International Conference on Urban Drainage (9ICUD), Portland, OR, USA, Sept. 8–13, 2002.
[https://doi.org/10.1061/40644\(2002\)251](https://doi.org/10.1061/40644(2002)251)
- [13] Rácz, T. "On the correction of processed archive rainfall data of siphoned rainfall", *IDŐJÁRÁS*, Accepted for publication September 2020. (in press)
- [14] Rácz, T. "On the correction of multiple minute sampling rainfall data of tipping bucket rainfall recorders", *IDŐJÁRÁS*, Accepted for publication April 2021. (in press)
- [15] OMSZ "Állomástörténet – Budapest" (Station history - Budapest), [online] Available at: https://www.met.hu/eghajlat/magyarorszag_eghajlata/eghajlati_adatsorok/Budapest/leirasok/tortenet/
- [16] Rácz, T. "The role of sampling and data processing in the reliability and comparability of precipitation maximum functions, in Hungarian", *Hidrológiai Közlöny*, 100(4), pp. 52–59, 2020. (in Hungarian)
- [17] Addinsoft "XLSTAT Statistical and data analysis solution", [online] Available at: <https://www.xlstat.com> [Access: 12 December 2020]
- [18] Costa, A. C., Soares, A. "Homogenization of Climate Data: Review and New Perspectives Using Geostatistics", *Mathematical Geosciences*, 41, pp. 291–305, 2009.
<https://doi.org/10.1007/s11004-008-9203-3>
- [19] Siswanto, S., van Oldenborgh, G. J., van der Schrier, G., Jilderda, R., van den Hurk, B. "Temperature, extreme precipitation, and diurnal rainfall changes in the urbanized Jakarta city during the past 130 years", *International Journal of Climatology*, 36(9), pp. 3207–3225, 2016.
<https://doi.org/10.1002/joc.4548>
- [20] Kotz, S. "Extreme Value Distributions. Theory and Applications", Imperial Collage Press, London, UK, 2000.
<https://doi.org/10.1142/p191>
- [21] Bozó, L. "Water and the atmospheric processes – atmospheric part of the hydrological cycle", *Magyar Tudomány*, 178(10), pp. 1198–1205, 2017. (in Hungarian)
<https://doi.org/10.1556/2065.178.2017.10.3>