Tipping bucket mechanical errors and their influence on rainfall statistics and extremes

P. La Barbera, L.G. Lanza and L. Stagi
Department of Environmental Engineering, University of Genova, 1 Montallegro, 16145 Genova, Italy

Water Science and Technology **Vol 45 No 2 pp 1–9** © IWA Publishing 2002

Abstract Based on the error figures obtained after laboratory tests over a wide set of operational rain gauges from the network of the Liguria region, the bias introduced by systematic mechanical errors of tipping bucket rain gauges in the estimation of return periods and other statistics of rainfall extremes is quantified. An equivalent sample size is defined as a simple index that can be easily employed by practitioner engineers to measure the influence of systematic mechanical errors on common hydrological practice and the derived hydraulic engineering design. A few consequences of the presented results are discussed, with reference to data set reconstruction issues and the risk of introducing artificial climate trends in the observed rain records.
Keywords Rainfall measurement; extremes, rain gauge; mechanical errors; calibration; statistics

Introduction

The measurement of liquid precipitation at the ground is crucial to most hydrological applications. Cumulated water depth over a given sampling interval in time is the usual output of traditional rain gauges, while inference of the rain intensity is possible by assuming uniform distribution of the rain process over the sampling interval. Such a measure of precipitation at the ground is usually assumed as the "ground truth" and newly developed techniques for rainfall monitoring based on remote sensing (weather radar, radiometers borne on aircrafts or satellite platforms, etc.) are commonly compared with the rain gauge figures for calibration and/or validation purposes.

As quoted by Marsalek (1981) the tipping-bucket rain gauge dates back to the 18th century and – though further refined over the years – it has become probably the most popular recording rain gauge, used by most national weather service agencies. The reason for such a widespread popularity comes from the very simple mechanics exploited for direct measurement of rainfall, and the reliability of the instrument. Also, it can be easily updated in its data acquisition and storage components as long as new electronic devices become operationally available. Finally, maintenance loads are reasonable and the cost is affordable even in the case of rather extended networks.

However many different types of water collectors have been implemented to convey rainfall from a standardised orifice into the measuring bucket and a large variability still exists world-wide in presently operational gauges. Simple gauges based on the tipping-bucket technology are also used for the measurement of other hydrological variables such as runoff and sediment transport.

The measurement of liquid precipitation at the ground is affected by different sources of both systematic and random errors, mainly due to wind, wetting and evaporation induced losses (e.g. Sevruk, 1982) which make the measurement of light to moderate rainfall scarcely reliable in the absence of an accurate calibration. Wind induced errors still have an influence at rainfall rates in the order of 20–50 mm/h with an incidence around 5% being reported at daily scale after comparison between paired ground level and elevated gauges in a few intercomparison stations in central Europe (Sevruk and Hamon, 1984). Solid precipitation measurements (snow) are even more difficult as snow is more sensitive than rain to weather related errors. Sampling errors due to the discrete nature of the rain measurement are also recognised to be dependent on the bucket size and sampling interval, though not on rain intensity, and can be analytically evaluated. Systematic errors are commonly accounted for in precipitation measurements by means of correction models that can be generally expressed in the form:

$$P_c = k\left[P_g + \quad {}_i \quad P_{gi}\right] \tag{1}$$

where P_c is the corrected figure, P_g is the gauge measured precipitation, ${}_i \quad P_{gi}$ is the sum of correction terms for various error sources, and k is the wind deformation coefficient. The detailed model, originally proposed by Sevruk (1982), was later modified by Legates and Willmott (1990) to account for both liquid and solid precipitation and can be written as:

$$P_c = k_r\left(P_{gr} + \quad P_{wr} + \quad P_{er} + \quad P_{mr}\right) + k_s\left(P_{gs} + \quad P_{ws} + \quad P_{es} + \quad P_{ms}\right) \tag{2}$$

where P_w, P_e and P_m are the correction terms for wetting, evaporation and mechanical errors respectively, while subscripts r and s refer to liquid (rain) and solid (snow) precipitation.

The tipping-bucket rain gauge is also known to underestimate rainfall at higher intensities because of the rainwater amount that is lost during the tipping movement of the bucket. The related biases are known as systematic mechanical errors and result in the overestimation of rainfall at lower intensities and underestimation at the higher rain rates. Though this inherent shortcoming can be easily remedied by dynamic calibration, usual operational practice in hydro-meteorological services and instrument manufacturing companies relies on single-point calibration, based on the assumption that dynamic calibration has little influence on the total recorded rainfall depth (Fankhauser, 1997). This results in the systematic underestimation of intense rainfall that can be quantified – e.g. in the SIAP family of operational gauges examined in this work – up to 10–15% at rain rates higher than 200 mm/h. Note that such intense rain rates may be commonly observed at very fine resolution in time even during precipitation events totalising low to intermediate intensities at the event scale. In case of intense events the extreme components of the intensity spectrum contribute to the core of the event, leading to higher average errors on the rain totals.

The relevant effect of this bias, which is shown to increase with rain intensity, is however much more evident on the derived rainfall statistics than it is with respect to the underestimation of rain totals (La Barbera *et al.*, 1998). Assessment of the return period of rainfall extremes, both at a single site and in the framework of regionalisation studies, is indeed significantly affected by systematic mechanical errors in case only single-point calibration of the gauge is performed. In this paper the influence of such errors over rainfall statistics and extremes is quantified based on the analysis of tipping-bucket rain gauges operated by the former National Hydrographic Service (NHS) of the Liguria region of Italy.

Mechanical errors and dynamic calibration

The very simple mechanics of the Tipping-Bucket Rain gauge (TBR) has been definitely understood for many years. Systematic mechanical errors due to the rainwater lost during the tipping movement of the bucket are also well-known and remediation by means of dynamic calibration has been proposed by several authors (see e.g. Calder and Kidd, 1978; Marsalek, 1981; Niemczynowicz, 1986).

In order to share a basis for quantitative estimation of the bias introduced on the statistics of extreme rainfall events that are derived from the elaboration of rain gauge records we will briefly recall in this section the origin of such errors and describe one common methodology for dynamic calibration of TBRs.

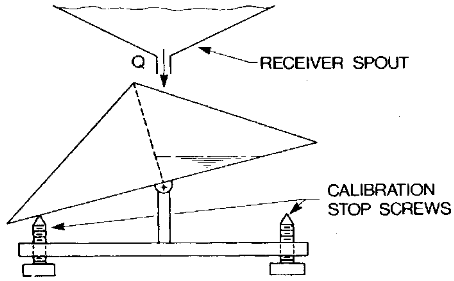

Figure 1 The tipping bucket mechanism (after Marsalek, 1981)

Water losses that are observed during the tipping movement of the bucket can be explained as follows. Consider the tipping movement (see Figure 1) as starting at that instant in time when the bucket is completely filled in with its nominal volume of water. Initiation of the movement of the two paired compartments is subject to inertial forces and completion of the bucket tipping around its rotation axis requires a certain amount of time. During such a time window the incoming rain intensity continues to supply water through the funnel. The amount of water received by the rotating bucket during half its tipping movement, i.e. during rotation until the compartment being emptied does not receive any more water from the receiver, is lost in the measurement process.

The relevance of such losses, affecting each single tipping of the bucket, increases with rainfall intensity and is a function of the total time T requested for the bucket to complete its rotation.

According to Marsalek (1981) the relationship between the recorded (I_r) and actual (I_a) intensities as a function of T is given by:

$$I_r \,/\, I_a = h_n \,/\, (h_n + I_a \lozenge\ T) \tag{3}$$

where h_n is the nominal rainfall depth increment per one tip. Note that $I_r / I_a = 1$ only in the case of $T = 0$ and that T is a function of rainfall intensity.

The relationship presented by Marsalek (1981, Figure 4) between T and I_a shows significant durations of the bucket movement ranging between 0.3 and 0.6 s for the instruments analysed. However, the uncertainty involved in the measurement of the time of tipping – due to the very slow initiation of the bucket rotation – made the comparison of experimental and theoretical calibration curves hardly appreciable in the author's work. Sophisticated measurement of T allows better success in comparing the experimental calibration curve with its theoretical expression.

However, direct estimation of the calibration curve is far more reliable than its theoretical derivation as it does not involve sophisticated measurements of very short intervals in time as a function of varying rain rates. A simple hydraulic apparatus can be used in this aim (see Figure 2b), which allows high precision measurements and reliable dynamic calibration of TBRs. The objective is that of providing the gauge receiver with a constant rain rate at a number of calibration points in the (I_a, I_r) space. This is achieved by connecting a constant water level tank with the receiver by interposition of a nozzle with specified diameter. By modifying the water head over the orifice and the nozzle diameter, constant flows can be generated at various flow rates as desired (see Lombardo and Stagi, 1997; Humphrey *et al.*, 1997).

Underestimation of intense rainfall

The estimation of the amount and characteristics of systematic mechanical errors was addressed in many previous works on calibration related issues (with bias quantified as about 10%) and early results were obtained by Becchi (1970) and Adami and Da Deppo (1985).

An extended survey of operational TBRs in use at the NHS in the Liguria region of Italy was preliminarily performed in this work by means of laboratory tests. The simple apparatus for dynamic calibration described in the previous section has been used for the assessment of systematic mechanical errors of some 60 instruments from various manufacturing companies. Forty out of the whole set of rain gauges analysed are currently used by the NHS, the others being operated by private companies or different organisations.

Calibration curves have been obtained for all the instruments analysed, according to a power law formulation that can be expressed as:

$$I_a = \quad \lozenge I_r \tag{4}$$

where I_a and I_r are again the actual and recorded rain intensities and , are the calibration parameters.

The associated relative measurement error can be defined as:

$$= \frac{I_r}{I_r - I_a} \lozenge 100 \tag{5}$$

so that < 0 indicates underestimation rather than overestimation (> 0) of the actual rain rates.

For the sake of conciseness only synthetic results from the aforementioned survey are reported here in order to quantify the relevance of systematic mechanical errors in operational gauges. All instruments have been carefully cleaned and verified for volumetric calibration in the laboratory before undergoing the dynamic calibration test. This ensures that malfunctions due to the blocking and ageing are avoided and the observed errors in recording of the prescribed rain rates are only due to the inherent mechanics of the TBR.

The dispersion of calibration parameters, grouped by type of rain gauge, can be observed by plotting the results in the (,) space. Relative measurement errors range from limited overestimation at the lower intensities (below the point of single calibration) to underestimation of as much as 10 to 15% at the highest rain rates. The result is well in agreement with the figures obtained in previous studies as briefly recalled in the opening of this section. The derived quantitative estimates of systematic mechanical errors can be therefore assumed as quite reliable and the related calibration curves (Figure 2a) can be used in the eventual assessment of the bias induced on the derived statistics of extreme events.

In this paper we will concentrate on one single family of TBRs among the various ones analysed, i.e. the one presenting the highest numerosity of the sample set. This allows more reliable evaluation of the average behaviour and estimation of the confidence we have in the derived average characteristics. The SIAP family of TBRs accomplishes such requirements for the Liguria region of Italy and is therefore analysed in more detail in this work.

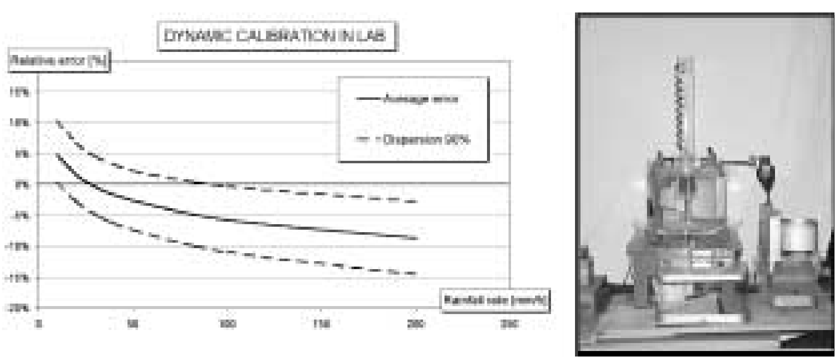

Figure 2 (a) Calibration curves for the analysed SIAP gauges and (b) calibration device

All numerical figures that are provided in the following therefore pertain to the “average SIAP gauge”, defined as the ideal gauge that behaves according to the average characteristics of a set of about 60 operational instruments sampled from that family.

The average calibration curve for the SIAP set of sample gauges is represented in Figure 2a, together with the 90% confidence interval on the estimation of the mean. The dispersion of individual values around the average behaviour is also indicated. At $I_a = 200$ mm/h we observe an average underestimation of about 8%, with a 90% chance of ranging between 2 and 15%.

Statistical influence of systematic errors

The main objective of the present work is to demonstrate that systematic mechanical errors – though affecting individual measures by a small percentage – induce significant biases in the results of the most common statistical analyses of rainfall extremes that are performed on uncalibrated datasets. These biases can be quantified, e.g. in the assessment of the return period T of short-duration/high-intensity rainfall, at about 100% for $T = 100$ years (see later in this section). The bias affects the expected value of the estimated return period, the amplitude of the associated confidence bands being not significantly modified.

The common statistical tools for the estimation of rainfall extremes have great relevance in many engineering applications (Kottegoda and Rosso, 1997) and are often used to derive design rainfall and flood figures following the reference values indicated by regulatory commitments. For instance, in the Liguria region of Italy practitioner engineers are required to design hydraulic structures interacting with the river flow according to design floods with $T = 200$ years. The limited number of observations that are generally available makes the estimation of the design flood only reliable within an associated confidence interval. Systematic mechanical errors of TBRs introduce a further source of uncertainty in the estimation of common statistics that we have quantified in the following through the definition of the “equivalent sample size”.

Our investigation starts with the analysis of the basic tools commonly used in the statistical analysis of rainfall extremes, i.e. with the exercise of fitting the available data series with a theoretical probability density function that is known (or expected) to reproduce the behaviour of the observed process. When dealing with extremes the simplest reference tool is the extreme value distribution known as the Gumbel distribution or EV1.

Systematic mechanical errors are assumed to hold in the form of the calibration curve expressed by Eq. (4), with and obtained from the set of operational gauges analysed in the previous section (all rain gauges were in operational use at the NHS of Liguria when tested in the laboratory). Since the relevant statistics are usually evaluated on the aggregated rain figures, at least over one hour in time, the first step is to assess the average error at that scale.

This was obtained by disaggregation of the hourly figures down to the resolution of five minutes, where the higher intensities are likely to occur and systematic mechanical errors are relevant. The disaggregation algorithm used is simply based on the statistics derived from the corresponding depth-duration-frequency curves, since the variance of the process is assumed to be the main factor controlling the results in terms of mechanical errors. In this aim, high resolution data from a single rain gauge managed by DIAM (Dept. Environmental Engineering) at the University of Genova has been used after aggregation of the original data (one minute) to the scale of 5 minutes. The internal distribution of rain intensities within the events was modelled based on such information. As the rain gauge mechanics is assumed to have no memory, disaggregation of hourly data was performed irrespective of the hyetograph shape, and the influence of the error on the disaggregated hourly data is shown in Figure 3a. Based on the derived hourly errors the Gumbel distribution was fitted to both the original and corrected data series, the two curves being plotted on

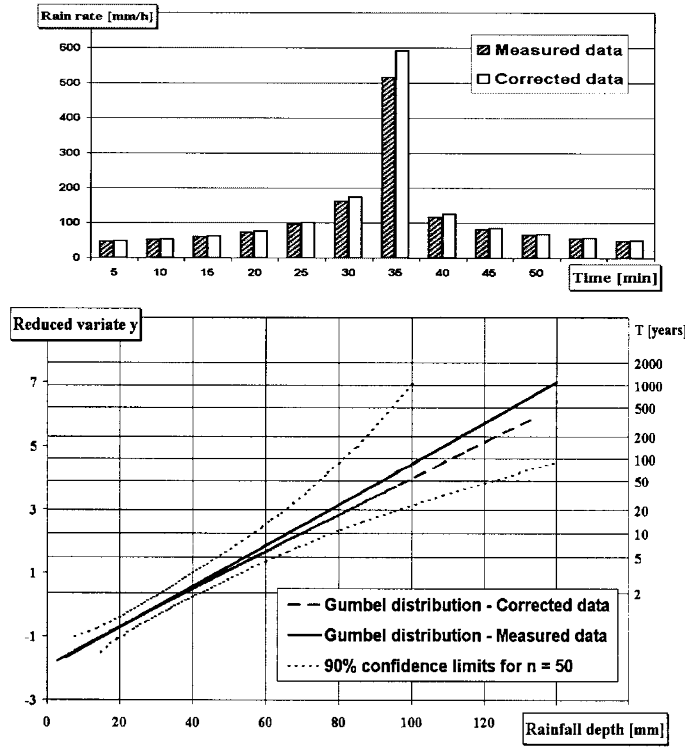

Figure 3 Statistical influence of systematic mechanical errors on the disaggregated rain data and the Gumbel distribution of rainfall extremes

the same graph of Figure 3b for better visualization of the derived bias on the statistics of extreme values.

In the hypothesis of an underlying Gumbel distribution of the extremes, the depth-duration-frequency curves for both the original and corrected data were derived, as reported in Figure 4a. In order to reduce probabilistic uncertainties due to the limited sample size we will also make reference here to the regionalisation studies performed on intense precipitation by one of the authors over the territory of the Liguria region of Italy (La Barbera, unpublished report). The sample size is made of about 2500 maximum yearly rainfall data over some hundred rain gauge stations operating, on average, for 25 to 30 years. With the hypotheses introduced above on the actual errors at the hourly scale the Two-Component Extreme Value (TCEV) distribution (Rossi *et al.*, 1984) was also fitted to the extended dataset, and the resulting curves are plotted in Figure 4b.

The overall influence of systematic mechanical errors on the overestimation of the return period associated with extreme rainfall events is represented in Figure 5 in terms of the ratio between the estimated (original data) and the actual (corrected data) return period. In the two graphs the ratios above are plotted against the duration of the rainfall and the corresponding area of the underlying hydrological basin whose response time is equal to the specified duration (assuming the simple one-to-one hydrological relationship $t = 0.2 + 0.27 \cdot A^{0.5}$ between the response time t and basin area A, which is calibrated for the basins of Liguria).

Finally, a quantitative measure of the uncertainty associated with relying on uncalibrated data sets for determination of rainfall statistics to be used in engineering applications has been derived. This is intended as a simple index that can be easily employed by practitioner engineers to measure the influence of systematic mechanical errors on common hydrological practice and the derived hydraulic engineering design. The index is

(a)

(b)

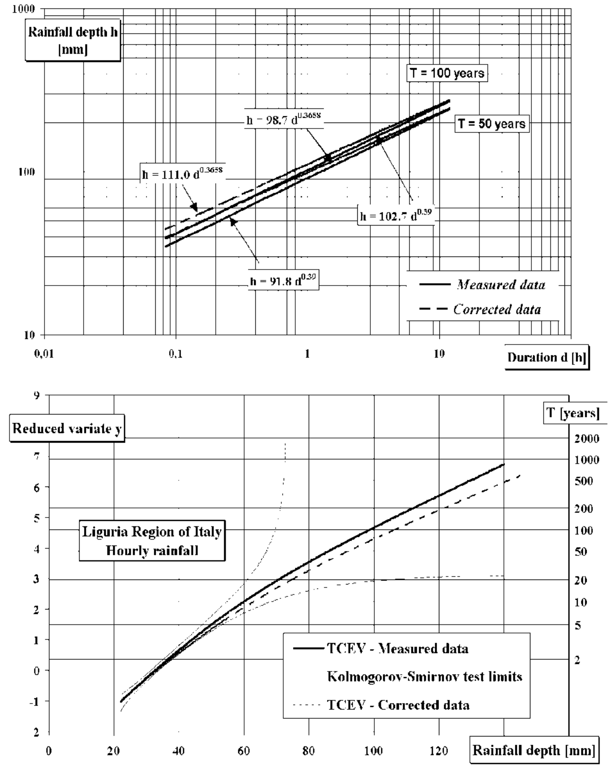

Figure 4 Statistical influence of systematic mechanical errors on the depth-duration-frequency curves (a) and the TCEV (Two-Component Extreme Value) distribution of rainfall extremes (Rossi *et al.*, 1984) (b)

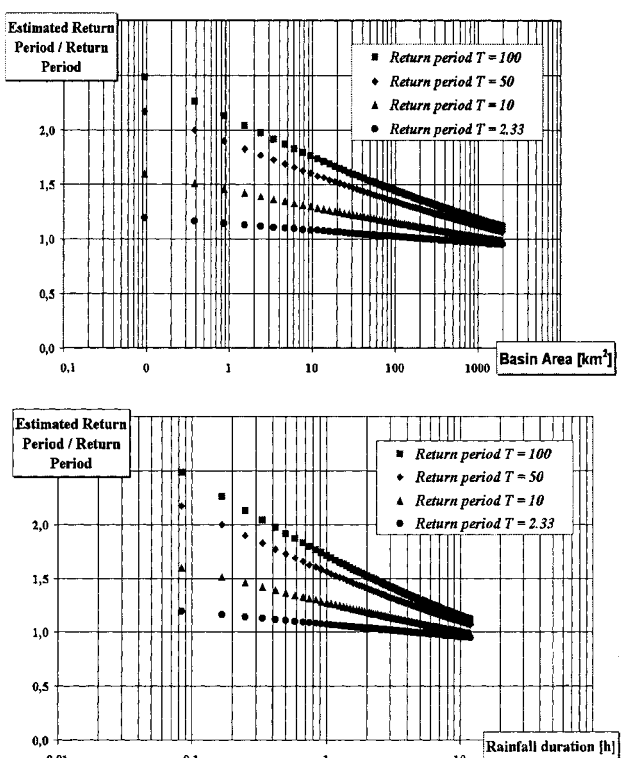

Figure 5 Statistical influence of systematic mechanical errors on the return period of extreme events

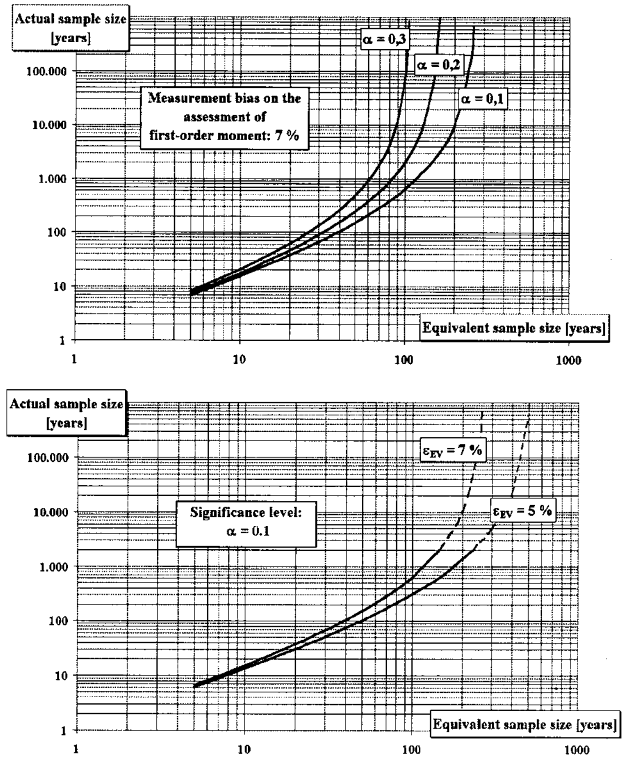

Figure 6 The equivalent sample size as a function of the mean error at aggregated scales $_{EV}$ and the confidence level on the assessment of the first order moment

expressed in the form of an "equivalent sample size", which quantifies the equivalent number of calibrated data that would be needed to achieve the same statistical uncertainty introduced by the influence of systematic mechanical errors in uncalibrated data sets. For the sake of simplicity, and in order to use quantities that are familiar to practitioner engineers, the uncertainty is here represented by the confidence intervals on the estimation on the mean over a set of extreme rainfall data (e.g. annual maxima) for a given duration. The duration is taken into account through the mean average error at aggregated scales, while confidence levels in the range from 0.1 to 0.3 are considered as typical values used in engineering design.

Since the limits of the confidence interval on the mean are defined as (see e.g. Kottegoda and Rosso, 1997):

$$P\ \ m-\frac{◊k\ _{/2}}{\sqrt{n}}<\quad <m+\frac{◊k\ _{/2}}{\sqrt{n}}\ =1- \tag{6}$$

where   is the theoretical (and unknown) expected value for the data set under consideration,   is the standard deviation (either theoretical or a sample statistic), $k\ _{/2}$ is the quantile such that $1-F_U(k\ _{/2})=\ /2$, n is the sample size, and m is the sample mean. Therefore the limits are sample statistics and random variables.

The mechanical error due to uncalibrated rain gauges can be suitably expressed as a function of the mean by resorting to the appropriate distribution function as:

$$e=e_{\%}(k_T)\,◊k_T\ \ ◊m \tag{7}$$

with k_T being the frequency factor (Chow, 1951). By assuming a specific probability distribution curve for the extreme values and for a given return period T, we can synthetically state that $e=\ \ \cdot m$.

By interpreting this error as a further source of uncertainty in the estimation of the mean, the e.g. upper limit of the confidence interval can be written as:

$$m+\frac{\Diamond k_{\ /2}}{\sqrt{n}}+\quad \Diamond m\ = m\ 1+cv\ \Diamond k_{\ /2}\ \Diamond\frac{1}{\sqrt{n}}\ +\ \ = m\ 1+\frac{cv\,\Diamond k_{\ /2}}{\sqrt{n_{EQ}}} \tag{8}$$

with cv being the coefficient of variation, and:

$$n_{EQ}=\frac{n}{(1+\quad /\ cv\,\Diamond k_{\ /2})^{2}} \tag{9}$$

is the desired equivalent sample size, which can be estimated for a given once is obtained by dynamic calibration, and a constant value is assumed for cv. In the graphs of Figure 6 the actual sample size n is plotted against the equivalent sample size n_{EQ} for different values of the confidence and the duration, the latter being basically expressed in terms of the coefficient .

Discussion and conclusions

Though very well known and easily avoidable by means of dynamic calibration, systematic mechanical errors of tipping-bucket rain gauges produce biases in the measurement of liquid precipitation that are commonly neglected in the management of many operational monitoring networks.

We have shown in this paper that such errors heavily affect the measurement of rainfall at higher intensities and the derived statistics, with non-negligible consequences on the numerical estimates of parameters involved in the common statistical tools that are used for characterisation of extreme events (GEV and TCEV distributions, depth-duration-frequency curves, etc.).

The bias introduced in the estimate of the return periods of extreme rainfall rates has been quantified in the case of single site and regionalisation studies based on figures from the analysis of the monitoring network operating in the Liguria region of Italy. Definition of the equivalent sample size allowed estimation of the role of systematic errors in reducing the advantage usually attributed to the high number of rain records that can be exploited for regionalisation of extreme rainfall and floods. This new index can be used by practitioner hydrologists in order to provide an estimation of the loss of significance associated with un-calibrated instruments, even in the presence of a high number of rain gauge stations. The equivalent sample size is innovative in this view, being the first quantitative index available to measure the actual reliability of rain series for statistical and design purposes, allowing even a revision of the present regulations about the return period of design rainfall to be used in hydraulic engineering design.

As a consequence of the above results one might conclude that dynamic calibration is compulsory for reliable rainfall monitoring based on traditional gauges, and weather service agencies should proceed in revising their instruments in order to obtain more accurate measurements for the future. Though this is certainly desirable it also raises non-trivial problems as regards the homogeneity of data series that are recorded before and after calibration of the rain gauges.

As mechanical errors affect the higher rain rates (especially those registered at very short intervals in time) recovering of past records and their correction is only possible at very fine resolution. Unfortunately, most of the historical information is derived by interpretation of the rain charts and anyway stored as accumulated rainfall values over intervals of 30 to 60 minutes at best. Details of the rain process at finer time scales are irremediably lost for a large number of measurement sites where up-to-date instruments are being progressively installed. Correction of past records in order to obtain homogeneous rainfall series can be however performed based on suitable disaggregation of the recorded figures, at least down to resolutions in the order of five minutes where the rain rate is higher and significant biases arise. This would allow only statistical correction and requires accurate investigation of the small scale structure of temporal rainfall, which may result site-dependent, with possible seasonality effects and climatic trends.

A huge amount of work is already available in the literature about disaggregation of point precipitation using various underlying models of the rain process. Results from sensitivity analysis over the various downscaling methodologies may be of some interest in this context, though it is already evident that the imposed variance of small scale rainfall will basically control – whatever the structure of the disaggregation model – the resulting correction of the statistical parameters.

One last comment about the homogeneity of temporal rainfall series relates to climate change. Indeed, any typical monitoring network that is continuously updated with more reliable gauges (as is presently the case for the Liguria region of Italy) would experience an artificial climatic trend towards increasing climatological precipitation in case mechanical errors affecting historical records are systematically neglected. Since rain gauge manufacturer companies are progressively distributing dynamically calibrated instruments, the risk of introducing artificial trends in rainfall series is far from being just academic.

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, 1–3 April 1985.

Becchi, I. (1970). Sulla possibilità di migliorare le misure pluviometriche: il controllo dei misuratori a vaschetta (in Italian). Technical Report for CNR Grant n. 69.01919. University of Genova, pp. 11.

Calder, I.R. and Kidd, C.H.R. (1978). A note on the dynamic calibration of tipping-bucket gauges. *J. Hydrology*, **39**, 383–386.

Chow, V.T. (1951). A general formula for hydrologic frequency analysis. *Trans. Am. Geophys.Un.*, **32**, 231–237.

Fankhauser, R. (1997). Measurement properties of tipping bucket rain gauges and their influence on urban runoff simulation. *Wat. Sci. Tech.*, **36**(8–9), 7–12.

Humphrey, M.D., Istok, J.D., Lee, J.Y., Hevesi, J.A. and Flint, A.L. (1997). A new method for automated calibration of tipping-bucket rain gauges. *J. Atmos. Oc. Techn.*, **14**, 1513–1519.

Kottegoda, N.T. and Rosso, R. (1997). *Statistics, Probability, and Reliability for Civil and Environmental Engineers*. McGraw-Hill, New York.

La Barbera, P., Lanza, L.G. and Stagi, L. (1998). Influence of measuring errors due to uncalibrated raingauges on flood analysis and prediction. (Abstract) EGS XXIII General Assembly, Nice, 20–24 April 1998. *Annales Geophysicae*, **15**(II), C-456, 1998.

Legates, D.R. and Willmott, C.J. (1990). Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatology*, **10**, 111–127.

Lombardo, F. and Stagi, L. (1997). Dynamic calibration of rain gauges in order to check errors due to heavy rain rates. *Proc. Int. Conf. On 'Water in the Mediterranean'*, Istanbul, 25–29 November. In press.

Marsalek, J. (1981). Calibration of the tipping bucket raingage. *J. Hydrology*, **53**, 343–354.

Maksimović, Č., Bužek, L. and Petrović, J. (1991). Corrections of rainfall data obtained by tipping bucket rain gauge. *Atmospheric Research*, **27**, 45–53.

Niemczynowicz, J. (1986). The dynamic calibration of tipping-bucket raingauges. *Nordic Hydrology*, **17**, 203–214.

Rossi, F., Fiorentino, M. and Versace, P. (1984). Two-component extreme value distribution for flood frequency analysis. *Wat. Resour. Res.*, **20**(7), 847–856.

Sevruk, B. (1982). Methods of correction for systematic error in point precipitation measurement for operational use. *Operational Hydrology Report No 21*, WMO Rep. No. 589, pp. 91.

Sevruk, B. and Hamon, W.R. (1984). International comparison of national precipitation gauges with a reference pit gauge. *Instruments and Observing Methods Report No. 17*, WMO, pp. 86.

Sevruk, B. and Klemm, S. (1989). Types of standard precipitation gauges. *Instruments and Observing Methods.* Proc. Int. Workshop on Precipitation Measurements. WMO Rep. No. 48, p. 227–232.