Phase space parametrization of rain: the inadequacy of the gamma distribution

Massimiliano Ignaccolo

Dept. of Earth & Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, North Carolina

Carlo De Michele

DIIAR, Politecnico di Milano, Milano, Italy

Abstract

We show that the Gamma distribution is not an adequate fit for the probability density function of drop diameters using the Kolmogorov-Smirnov goodness of fit test. We propose a different parametrization of drop size distributions, which not depending by any particular functional form, is based on the adoption of standardized central moments. The first three standardized central moments are sufficient to characterize the distribution of drop diamters at the ground. These parameters together with the drop count form a 4-tuple which fully describe the variability of the drop size distributions. The Cartesian product of this 4-tuple of parameters is the rainfall phase space. Using disdrometer data from 10 different locations we identify invariant, not depending on location, properties of the rainfall phenomenon.

1 Introduction

At "punctual" space scale ($\sim$50 cm^2) rain can be described by a stochastic sequence of couples (D_j, τ_j): D_j being the diameter of the j-th drops, and τ_j the interval of time between the arrival of the j-th drop and the $(j+1)$-th drop $(j = 1, 2, 3, ...)$. Partitioning the time axis in sampling time

intervals of equal duration it is possible to build from the sequence (D_j, τ_j) the sequence $(p_k(D), N_k)$, with $p_k(D)$ the probability density function of drop diameter, and N_k the number of drops observed at the ground during the l-th sampling time interval ($l = 1, 2, ...$). These two last quantities are usually measured by disdrometers with a sampling time in the range 10s-5min. Observational evidences suggest that rain is a non-homogeneous process as the variability observed in the sequence $(p_l(D), N_l)$ cannot simply ascribed to the variability expected when sampling from a single stochastic process [21, 14, 7, 5, 6]. The sequence (D_j, τ_j), and thus $(p_l(D), N_l)$, is non-stationary: the "statistical rules" to which the couples (D_j, τ_j) obey are not invariant under time translation [24].

A proper description of the distribution of drop sizes, which is fundamental for understanding the microphysics involved in the mechanisms of precipitation formation and retrieving rainfall from radar sensors, has been the main research goal since the introduction of the impact disdrometer has made reliable and prolonged measurements of drop diameters feasible. The main tool the meteorological community uses to describe the variability of the rainfall phenomenon is the drop size distribution, namely the concentration

$\mathcal{N}(D)$ per unit volume and diameter. In their seminal work Marshall and Palmer [15] proposed the exponential as functional dependence for $\mathcal{N}(D)$ with a decaying constant depending on the rainfall rate. Later, Joss and Gori [9] made clear that the exponential form is the results of sampling long time intervals (∼30 minutes), and at smaller time intervals (∼1 minute) the drop size distribution does not have an exponential functional dependence. Ulbrich [25] proposed the Gamma distribution as the functional form for the drop size distributions sampled at short time intervals. The rationale for this choice is that while an exponential decay is observed for the right tail of the distribution (recently [26] have shown that break-up due to air viscosity is the main mechanism underlying the occurrence of exponential right tail), the density at smaller diameters is smaller than that of an exponential distribution. The Gamma distribution has been the only choice adopted in literature (if one excludes sporadic attempts with the log-Normal distribution, see e.g. [3], [19], and [17]) mainly because of two properties: 1) it is defined by two parameters, and 2) its moments can easily calculated so that any rain bulk variable can easily expressed in terms of the parameters of the Gamma distribution. The Gamma distribution has been widely accepted by radar meteorology and cloud physics communities, even if measurements show that the Gamma distribution is not general enough to represent ade-

quately the full range of the sample variability, [11]. The fit accuracy of the Gamma distribution to disdrometer data has not been properly addressed (aside from subjective statements: "a good fit" or "a reasonable fit" not supported by any objective measurement) or mistakenly addressed (see Section 4) as in [20]. In the present work, we consider data from 10 different sites, and use the Kolmogorov-Smirnov goodness of fit test (e.g. [10, 12]) to show that the Gamma distribution is a poor fit to 1 minute sampled drop size distribution.

To obviate to this inadequacy, we propose a parametrization of the drop size distribution based on a standard procedure in statistical science: the description of an unknown probability density function in terms of its mean μ, standard deviation σ, skewness γ, and kurtosis κ. In particular, we show that mean, standard deviation and skewness are the minimum number of parameters, which effectively describe the drop size distributions. These parameters, together with the drop count N, are the variables necessary to describe the rainfall phenomenon at a punctual scale in space and at short time scales. All bulk variables of interest can be derived by the 4-tuple of parameters (N, μ, σ, γ). Adopting the jargon of the scientific community studying dynamical systems, we refer to the four dimensional Cartesian product $\mathbf{R}^4$ spanned by the 4-tuple (N, μ, σ, γ) as the rainfall phase space.

The work is organized as follows. In Section 2, we discuss the data used in our analysis, Section 3 we illustrate all the methods of analysis adopted. Section 4 reports our results, and Section 5 our conclusions.

2 Data

We consider Joss-Waldvogel disdrometer data sampled at 1 minute time intervals from ten different locations on Earth's surface. Table 1 gives the list of the locations with a three letter code used to reference the site in the Figures. In addition, the Kppen-Geiger climate classification ([13]) of each site is provided. Table 2 completes the description of the data sets providing for each site the latitude, longitude, altitude, together with number of minutes and drops considered in the analysis.

2.1 Data processing

All data are processed as follows. We consider for each data base only minutes for which the drop count is $\geq$60 (drop arrival rate $\sim$1 per second). The rationale for this choice is twofold: A) It guarantees a minimum reasonable accuracy for the estimation of the probability density function $p(D)$, and the other statistical parameters (mean, standard deviation, ...). B) It excludes time intervals of observation which are quiescent (sparse precipitation) whose contribution to the total cumulated precipitated volume is negligible (see discussion in [7]). After this first filtering, we eliminate from the remaining minutes of observation any outlier drop count. For each minute, we find the disdrometer class with the maximum probability density, and we calculate the central continuous non-zero span of the probability density: the set of contiguous disdrometer classes with non zero counts and which includes the class with maximum density. The counts in all disdrometer classes which do not belong to the central continuous non-zero span are considered as outliers and are discarded: e.g. the disdrometer count (3,11,18,31,30,35,80,52,41,39,44,21,5,0,1,0,0,0,0,0) has an outlier in the 15-th class which is disregarded leading to the count (3,11,18,31,30,35,80,52,41,39,44,21,
5,0,0,0,0,0,0,0). The removal of outliers drop counts improves the estimate of higher moments of the probability density function $p(D)$. A detailed discussion on outliers and their effects on estimated statistical parameters can be found in [5].

3 Methods

We now describe the methods used to quantify the variability of the rainfall phenomenon.

3.1 Variability of drop diameters

Two equivalent descriptions are possible for the variability of drop diameters. 1) The drop size distribution defined as the concentration per unit volume and unit diameter $\mathcal{N}(D)$

$$\mathcal{N}(D) = N_V f(D), \tag{1}$$

where N_V is the number of drops per unit volume and $f(D)$ is the probability density function of drop diameter in the unit volume. 2) The flux-equivalent of Eq.(1): $N_V \to N$ and $f(D) \to p(D)$. N is the number of drops observed at the ground, and $p(D)$ is the probability density function of drop diameter at the ground. The two descriptions are equivalent as

$$p(D) = \frac{A_m T v(D)}{N} N_V f(D), \tag{2}$$

and

$$N = A_m T N_V \int_0^\infty v(D) f(D) dD. \tag{3}$$

In the above equations, T is the time interval of observation (in seconds), A_m the capture area of the instrument (in m^2), and $v(D)$ the drop velocity of diameter D (in m/s). Usually it is assumed that the arrival velocity of drops is equal to their limit velocity and $v(D) = CD^b$, where $C = 3.78$ m/(s $\mathrm{mm}^{0.67}$) and $b = 0.67$ ([25]). With these definitions

$$p(D) = \Xi D^{0.67} N_V f(D) \qquad \Xi = \frac{A_m T C}{N}. \tag{4}$$

By use of Eq.(4), we can connect the moments $M_{\alpha,p}$ of the probability density function observed at the ground with those of the concentration per unit volume $M_{\alpha,\mathcal{N}}$:

$$M_{\alpha,p} = \Xi M_{(\alpha+0.67),\mathcal{N}} \Longleftrightarrow M_{\alpha,\mathcal{N}} = \Xi^{-1} M_{(\alpha-0.67),p}. \tag{5}$$

Using the above equation one is able to derive expressions for any rain bulk variable. E.g. the rainfall rate R (in mm/h) is

$$R = 6\pi 10^{-4} C M_{3.67,\mathcal{N}} = \frac{6\pi 10^{-4}}{A_m T} N M_{3,p}. \tag{6}$$

The concentration per unit volume, Eq.(1) is by far the most common quantitiy in literature, even if measurements of drop sizes at the ground are by large the only available data. Hereby, we adopt the probability density of drop diameter at the ground, Eq.(2), as measured by disdrometer counts.

3.2 The Gamma distribution for $\mathcal{N}(D)$ and $p(D)$

The most common functional form adopted for the probability density function $f(D)$ in Eq.(1) is the Gamma distribution

$$f(D) = f_\Gamma(D, \lambda, k) = \frac{\lambda^{k+1}}{\Gamma(k+1)} D^k \exp(-\lambda D), \tag{7}$$

where k is the shape,λ the inverse scale parameter, and $\Gamma(x)$ is Gamma function. Consequently the statistical moment of order α of the drop size distribution $\mathcal{N}(D)$ is

$$M_{\alpha,\mathcal{N}} = N_V \int_0^\infty dD D^\alpha f_\Gamma(D, \lambda, k) = N_V \frac{\Gamma(k+\alpha+1)}{\Gamma(k+1)} \lambda^{-\alpha}. \tag{8}$$

However, it is common practice, in Literature, to write Eqs.(1) and (7) as

$$\mathcal{N}(D) = N_0 D^k \exp(-\lambda D), \tag{9}$$

where N_0 is defined as

$$N_0 = N_V \frac{\lambda^{(k+1)}}{\Gamma(k+1)}. \tag{10}$$

With the above notation

$$M_{\alpha,\mathcal{N}} = N_0 \Gamma(k+\alpha+1) \lambda^{-(k+\alpha+1)}. \tag{11}$$

Finally, if the probability density function $f(D)$ is a Gamma distribution, then also the probability density at the ground $p(D)$ is a Gamma distribution

$$\mathcal{N}(D) = N_v f_\Gamma(D, \lambda, k) \Longleftrightarrow p(D) = f_\Gamma(D, \lambda, k + 0.67). \tag{12}$$

3.3 Fitting the Gamma distribution to $\mathcal{N}(D)$ and $p(D)$

We briefly review the two main methods which have been adopted in literature to obtain the three parameters (N_V, λ, k) in Eqs.(1) and (7).

3.3.1 Method of the Moments (MM)

The method of moments (MM) uses the moments of the observed distribution to derive the parameters of the desired fitting function. For a drop size distribution with a Gamma distribution for diameter density, the $\mathrm{MM}_{\mathrm{n1,n2,n3}}$ method finds the number of drops per unit volume N_V, the scale k and shape λ such that the resulting distribution $\mathcal{N}(D)$ exactly matches the moments of order n1, n2 and n3 of the observed distribution. Hereby, we use the $\mathrm{MM}_{3,4,6}$ and $\mathrm{MM}_{2,3,4}$ procedures as they are the ones usually adopted in Literature (e.g. [23] and [22]), together with the $\mathrm{MM}_{0.67,1.67,2.67}$ adopted by [20]. This last procedure finds the concentration N_V, the scale k and shape λ of the drop size distribution $\mathcal{N}(D)$ matching the observed number of drops at the ground N (moment 0.67), the observed average drop diameter at the ground μ (moment 1.67), and the observed second moment of drop diameter at the ground $M_{2,p}$ (moment 2.67). Note that the last two conditions imply that the resulting drop size distribution matches the standard deviation of drop diameter σ observed at the ground. Therefore, we will use for brevity the notation $\mathrm{MM}_{N,\mu,\sigma}$ instead of $\mathrm{MM}_{0.67,1.67,2.67}$.

Using Eqs. (1) and (8), the parameters N_V, λ, and k are

$$\mathrm{MM}_{3,4,6}\begin{cases} N_v &= \frac{\Gamma(k+1}{\Gamma(k+5)}\lambda^4 M_4 \\ \lambda &= (k+4)\frac{M_3}{M_4} \\ k &= \frac{11G-8+\sqrt{G^2+8G}}{2-2G} \quad G = \frac{M_4^3}{M_3^2 M_6} \end{cases} \tag{13}$$

$$\mathrm{MM}_{2,3,4}\begin{cases} N_v &= \frac{\Gamma(k+1)}{\Gamma(k+2.67)}\lambda^{1.67} M_{1.67} \\ \lambda &= (k+3)\frac{M_2}{M_3} \\ k &= \frac{-7G+\sqrt{G^2+4G}}{2} \quad G = \frac{M_3^2}{M_2 M_4} \end{cases} \tag{14}$$

$$\mathrm{MM}_{\mathrm{N,\mu,\sigma}}\begin{cases} N_v &= \frac{\Gamma(k+1)}{\Gamma(k+2.67)}\lambda^{1.67} M_{1.67} \\ \lambda &= (k+1.67)\frac{M_{0,67}}{M_{1.67}} \\ k &= -4.34G + \sqrt{G^2+4G} \quad G = \frac{M_{1.67}^2}{M_{0.67} M_{2.67}} \end{cases} \tag{15}$$

3.3.2 Method of Maximum Likelihood (MML)

The method of maximum likelihood is the main alternative to the method of moments. Let $f_X(x,\theta)$ be the probability density function of the variable X given the vector of parameters θ, and $x_1,\ x, ...,\ x_N$ is a sample of size N. The likelihood $L(\theta)$ that the sample $(x_1, x_2, , ..., x_n)$ is drawn from the distribution $f_X(x,\theta)$ is defined as $L(\theta) = \prod_{i=1}^{N} f_X(x_i,\theta)$. In practice it is more convenient to deal with the logarithm of the likelihood, denominated the log-likelihood $\ln L(\theta) = \sum_{i=1}^{N} \ln f_X(x_i,\theta)$, or the average log-likelihood $l(\theta) = \frac{1}{N}\ln L(\theta)$. The ML method makes an estimation of θ maximizing the average log-likelihood, i.e. $\theta_{MML} = arg\max_{\theta}[l(\theta)]$, [2]. Dealing with disdrometer data, it is important to recall that disdrometers collect drops with a diameter $D \geq D_{min} = 0.3$ mm (for the JW disdrometer), making a lower truncation in the sample distribution. In the absence of small drops in sample datasets, the method of maximum likelihood ignoring this problem exhibits large bias which do not decrease increasing the sample size, [11]. Consequently, modifications to the MML are necessary to deal explicitly situations where lower truncations to the samples are present. We will consider the lower truncated Gamma distribution which density is

$$f(D) = f_\Gamma(D, D_{min}, \lambda, k) = \frac{\frac{\lambda^{k+1}}{\Gamma(k+1)} D^k \exp(-\lambda D)}{1 - \frac{\gamma(k+1,\lambda D_{min})}{\Gamma(k+1)}},\ D > D_{min} \qquad (16)$$

where $\gamma(k+1, \lambda D_{min})$ is the incomplete Gamma function calculated in λD_{min}. According to [8], the average log-likelihood is

$$l(\lambda,k) = -\ln\left(1 - \frac{\gamma(k+1,\lambda D_{min})}{\Gamma(k+1)}\right) + (k+1)\ln\lambda - \ln\Gamma(k+1) + \\ +k\left(\frac{1}{N}\sum_{i=1}^{N}\ln D_i\right) - \lambda\left(\frac{1}{N}\sum_{i=1}^{N} D_i\right). \quad (17)$$

The estimates of λ and k are obtained numerically minimizing the function $l(\lambda, k)$, using the R code provided by [8] in their appendix.

3.4 Statistical characterization of a probability distribution function

Under some general condition (e.g. [18]), a probability density function is completely determined by its moments: given the sequence of moments $\{M_j\}$, $j = 1, 2, 3, \ldots$ there exists an unique $f(x)$ such that $M_j = \int x^j f(x) dx$. This fact has lead to the moment-characterization, in statistical sciences, of probability density function for which a known parametric form is not available. For this purpose a suitable number of moments and/or function of moments will provide information on the unknown probability density function $f(x)$ and, as a consequence, on the dynamical process driving the realizations of the stochastic variable x. We refer to these parameters as the *statistical descriptors* of the probability density function.

The two most commonly used statistical descriptors are the mean $\mu = M_1$, and the standard deviation $\sigma = \sqrt{M_2 - (M_1)^2}$. In addition to these two parameters, the skewness γ, measuring the asymmetry of the distribution, and the kurtosis κ, measuring the peakedness of the distribution, are used. Skewness and kurtosis are the third and the fourth standardized central moments (the expectation value of $[(x-\mu)/\sigma]^3$ and $[(x-\mu)/\sigma]^4$) and can be written in terms of the moments of the distribution as follows:

$$\gamma = \frac{M_3 + 2(M_1)^3 - 3M_1 M_2}{[M_2 - (M_1)^2]^{3/2}} \tag{18}$$

and

$$\kappa = \frac{M_4 - 3(M_1)^4 + 6(M_1)^2 M_2 - 4M_1 M_3}{[M_2 - (M_1)^2]^2}. \tag{19}$$

Higher standardized moments do not have a particular name and are generally not used since the higher the moment the larger are the inaccuracies of any estimate. However, hereby we will make use of the fifth standardized central moment, the expectation values of $[(x-\mu)/\sigma]^5$, which will denote with the letter η

$$\eta = \frac{M_5 + 4(M_1)^5 + 10(M_1)^2 M_3 - 10(M_1)^3 M_2 - 5M_4 M_1}{[M_2 - (M_1)^2]^{5/2}}. \tag{20}$$

Mean, standard deviation, skewness, and kurtosis are sufficient to achieve a satisfactory (albeit not full) description of any probability density function of interest. In many cases they are redundant as some of the statistical

descriptors are shown to be function of the others. For example if $f(x) = f_\Gamma(x, \lambda, k)$ then only two elements of the 4-tuple $(\mu, \sigma, \gamma, \kappa)$ are independent. This redundancy also occurs for rainfall as we will show in Section 4.

3.4.1 Phase space

For a dynamical system, the term "phase space" indicates the Cartesian product $\mathbf{R}^n$ of the n variables necessary to describe the system. Note that the cardinality (n) of a dynamical system is always larger or equal to its degrees of freedom. To describe the motion a point particle of mass m in one dimension we need its position x and its quantity of motion p ($p = mv$, with v =velocity). In this case the phase space is the xp plane ($\mathbf{R}^2$), and at each time t the state of the particle is associated to a point in the xp plane. The time evolution of a dynamical system is described by the time evolution (trajectory) of its associated point in the phase space.

In the case of the rainfall phenomenon, the variables of interest are bulk variables (e.g rainfall rate, reflectivity), which implies a "summation" over the number of drops in a given interval of time. All the bulk variables are functions of either the couple $(N_V, f(D))$, if one uses the concentration per unit volume, or the couple $(N, p(D))$, if one uses the flux-equivalent description. If the probability density function at the ground $p(D)$ (or that in a unitary volume $f(D)$) has a parametric description: $p(D) = p(D, \theta_1, \theta_2, ..., \theta_m)$ in term of m parameters, then we define the phase space of rainfall as the Cartesian product $\mathbf{R}^{m+1}$ spanned by (m+1)-tuple $(N, \theta_1, \theta_2, ..., \theta_m)$. E.g., if we consider the concentration $\mathcal{N}(D)$ and assume a Gamma distribution for the probability density function $f(D)$ in Eq.(1), the rainfall phase space is $\mathbf{R}^3$ spanned by the 3-tuple (N_V, λ, k). Hereby, we do not impose any particular functional form on the function $p(D)$ (and thus $f(D)$). We show (Section 4) that the parameters μ, σ, and γ are sufficient to describe the variability of the probability density function at the ground $p(D)$. Therefore the Cartesian product $\mathbf{R}^4$ spanned by the 4-tuple (N, μ, σ, γ) can be considered as the rainfall phase space.

Any bulk variable B can be written as function of the phase space parameters $B = B(N, \mu, \sigma, \gamma)$. In the case of the rainfall rate R (expressed in mm/h)

$$R = \frac{6\pi 10^{-4}}{A_m T} N \left[\mu^3 + 3\mu\sigma^2 + \sigma^3\gamma \right]. \tag{21}$$

This equation states that the set of observation time intervals (1 minute in our case) with equal rainfall rate R is a three dimensional manifold in the phase space.

3.5 Kolmogorov-Smirnov's goodness fit for probability distribution functions

The Kolmogorov-Smirnov (K-S) goodness-of-fit test is a non-parametric test used to check if a sequence of random samples can be considered as a realization of a stochastic process with a given cumulative distribution function $F(x)$. The test compares the hypothetical $F(x)$ with the cumulative frequency $F_N(x)$, where $F_N(x) = i/(N+1)$ for $x_{(i)} \leq x < x_{(i+1)}$, $x_{(i)}$ is the i-th order statistics, and $i = 1, .., N$. The K-S uses as test statistic the maximum difference $D_N = max|F(x) - F_N(x)|$. If no parameter in $F(x)$ is determined from data, then D_N has a distribution which is independent by $F(x)$. Thus the critical value of D_N for a significance level of 5% and for large samples, $N > 35$, is $1.3581/\sqrt{N}$, and reported in all statistical textbooks (see e.g. [12]). Contrary, if the parameters of $F(x)$ are estimated, then the distribution of D_N is dependent on $F(x)$, and the critical value of D_N must be re-calculated, e.g. via Montecarlo simulations ([10]). The critical value of D_N re-calculated is always smaller than the value corresponding to the canonical case where it is assumed that "no parameter in $F(x)$ is determined from data". Disdrometer data report the occurrence of a drop in a given range of diameter values (diameter class) and not an "exact" diameter value which is needed to perform the Kolmogrov-Smirnov test. To bypass this limitation we assign to a drop in the j-th diameter class a random value selected uniformly in the range defined by the class itself.

4 Results

All the results reported in this Section refer to probability density function observed at the ground $p(D)$. Similar results can be obtained for the concentration per unit volume and diameter $\mathcal{N}(D)$ since the two distribution are connected via Eq.(2).

4.1 Measuring the adequacy of the Gamma distribution

The method of the moments $\mathrm{MM}_{\mathrm{n1,n2,n3}}$ for the Gamma distribution finds the number of drops per unit volume N_V, the scale k and shape λ such that the fitting concentration per unit volume and unit diameter exactly matches the moments of order n1, n2 and n3 of the observed concentration per unit volume and unit diameter $\mathcal{N}(D)$. None of the $\mathrm{MM}_{2,3,4}$, $\mathrm{MM}_{3,4,6}$, and $\mathrm{MM}_{N,\mu,\sigma}$ methods exactly match the moment 3.67, so that the "reproducibility" of the rainfall rate (observed versus the one derived from the fitted parameters) has been considered as a "measure" of fit goodness: e.g. ([23]) and [20]). Hereby , we show that accuracy with which the $\mathrm{MM}_{\mathrm{n1,n2,n3}}$ method reproduces the j-th moment of the concentration $\mathcal{N}(D)$ cannot be taken as a measure of fit goodness as the accuracy depends on the separation between j and the orders $n1$,$n2$,$n3$ and not just on the particular functional form chosen.

Let us consider the $\mathrm{MM}_{3,4,6}$ method. The fitting distribution by construction matches the third, fourth and sixth moment of $\mathcal{N}(D)$. Therefore it is not surprising that the 3.67-th moment of the fitting $\mathcal{N}(D)$ (the rainfall rate) is close to the observed value: the middle-left panel of Figure 1 indicates the relative error χ is bounded in the range -0.5%,0.5%. However, instead of a Gamma distribution for the functional form of $f(D)$, one could use any other distribution with two parameters (e.g. Gaussian, Lognormal, Beta) and obtain similar accuracies. What about the number of drops observed at the ground N (the 0.67-th moment of $\mathcal{N}(D)$)? As shown in the middle-right panel of Figure 1, the agreement is not so good as relative error of the order of $\pm$25% are possible. Next, we consider the $\mathrm{MM}_{N,\mu,\sigma}$ method. It reproduces the rainfall rate reasonably well, relative error bounded in the -5%,5% range (upper-left panel of Figure 1), but poorly reproduces the reflectivity Z (the sixth moment of $\mathcal{N}(D)$) as relative errors larger than $\pm$25% are not so uncommon (upper-right panel of Figure 1). Finally, the $\mathrm{MM}_{2,3,4}$ method reproduces with the same accuracy of the $\mathrm{MM}_{3.4.6}$ method (extremely well) the rainfall rate, the relative error is bounded in the range -0.5%,0.5% (bottom left panel of Figure 1), but is better than the $\mathrm{MM}_{3.4.6}$ method with respect the drop count N (2 is closer to 0.67 than 3), relative error in the range -15%,15% (bottom right panel of Figure 1).

Figure 1 suggests that the Gamma distribution, as convenient and as parsimonious it may be, is not a satisfactory functional form for the drop size distribution. To prove this point we use a proper measure of goodness-

of-fit, such as the Kolmogorov-Smirnov test, and what is considered to be the best fitting procedure, the MML method. In Table 2, we report the percentage of acceptance (ACP) and rejection (RJC) of the lower truncated Gamma distribution with parameters estimated using the MML, using the Kolmogorov-Smirnov goodness-of-fit test with a 5% level of significance to each minute of the DRW data (6863). In columns 2 and 3, the percentages are calculated using as 5% critical value $1.3581/\sqrt{N}$, the classical value reported in all statistical textbooks (e.g. [12]) assuming that no parameters of the Gamma distribution are estimated. In columns 4 and 5 the critical value is determined via Montecarlo simulations taking into account the fact that the parameters are estimated via the MML from data. The first row (ALL) reports the fraction of the total number of minutes in the DRW data sets for which the Gamma distribution can be considered a good fit. The percentage of acceptance passes from 71% to 45% when one takes in proper consideration that the parameters of the distribution are obtained from the sample ([10]). The remaining rows report the percentage of acceptance and rejection for subsets obtained using as thresholds the percentiles of the distribution of number of drops N per minute: 5%-, 25%-, 50%-, 75%-, 95%-percentile. E.g. the second row reports the results for the subsets with number of drops smaller than the 5%-percentile ($N_{5\%} = 66$ in our case): minutes with a small sample size. On the other hand, the last row reports the results for the subsets with number of drops larger than or equal to the 95%-percentile ($N_{95\%} = 1589$ in our case): minutes with a large sample size . We see that when we move to subsets of minutes with large drop counts the percentage of acceptance (rejection) diminishes (increases).

In summary, if we consider the entire DRW data set, we are confident (at the 95% level) that the Gamma distribution can be a proper fit for probability density function $p(D)$ only for 45% of 1 minute sampling time intervals. More disturbingly the percentage of rejection increases as the sample size increases. Note that test like the Kolmogorov-Smirnov should be administered to sample with a size of at least $\sim$100 ([12]) to have of any significance (in other words if the sample size is very small the effect of random fluctuations is large enough that almost any tested distribution will pass the test). Results for the other nine data sets (no reported here for brevity) are similar to that of Darwin. On the ground of these results, we reject the Gamma distribution as a proper fit for drop size distributions.

4.2 Rainfall phase space

We show that the value of standardized central moments of order ≥ 4 are strictly dependent from than the third one (skewness). In particular we study the dependence on the skewness γ of the fourth standardized central moment (kurtosis κ) and the fifth one denoted by the symbol η. As a consequence, mean (μ), standard deviation (σ), and skewness (γ) provide a satisfactory description of the variability of the probability density function $p(D)$ and therefore the 4-tuple (N, μ, σ, γ) can be considered as the rainfall phase space.

4.2.1 higher standardized central moments

To examine the dependence of the parameters κ and η on γ, we calculate for each data set the median, 5%-, and 95%-percentile of the observed values of κ and η for a given value of γ (in practice this is accomplished dividing each data sets in subsets with "equal" (± 0.08) value of skewness). The results are reported in Figure 2. The median values are depicted with solid lines of different colors one for each dataset. We also calculate, merging all the datasets, the 5% and 95% percentile for any given range value of the skewness if at least 100 samples (1 minute time interval of observation) are present. The range between the 5% and 95% percentile is shaded in gray in the figure. We see how in both cases (η vs γ, central panel and κ vs γ lower panel) the median lines are independent from the site chosen. The discrepancies observed for values of skewness larger than $\sim$2.56 are mostly due to lack of statistics as shown in the upper panel of Figure 2, where we plot the number of sample M for each range value of the skewness.

4.2.2 Relationship between phase space parameters

Figure 2 shows that mean μ, standard deviation σ and skewness γ are effective statistical descriptors for the probability density function of drop diameters at the ground $p(D)$. Next, we show that these three parameters are not independent. We divide the range of value of the skewness in intervals of length 0.64 ([-1.60,-0.96],...,[2.88,3.52] as in [5, 6]), and the values of the mean diameter μ as follows: [0.3,0.4], [0.4,0.5], [0.5,0.6], [0.7,0.8], [0.9,1], [1,1.2], [1.2,1.4], [1.4,1.6], [1.6,1.8], [1.8,2.0], [2.0,2.5], and [2.5,3] (indicated by the vertical dashed lines in Figure 3). The rationale for these choices is to have inside each range of skewness and mean a "reasonable" number (¿10) of 1 minute interval observations to calculate the median value of the

parameter σ. The results are reported in Figure 3. We see how for value of the mean diameter less than 1 mm the median curves are approximatively linear and do not depend on the site of observation. In this case, the site average angular coefficient and average intercept are reported at the bottom right of each panel, if at least the median from 5 sites was available. We see how the slope (intercept) increases (decreases) with increasingly larger value of the skewness until a sort of plateau is reached for the skewness ranges $[1.60, 2.24]$ and $[2.24, 2.88]$ after which the slope decreases (intercept increases) again. For values of the mean diameter larger than 1 the median curves depend on the particular site of observation. However, the median curve is also estimated from a smaller number of samples (in the range10-100). Thus, with the present data, we cannot judge if the discrepancies between sites for the median curve (in the range $\mu > 1$) are real properties of the rainfall phenomenon or merely artifacts due to the poor statistical sampling available.

In summary, if μ is in the interval [0.3,1] then

$$\sigma = a(\gamma) + b(\gamma)\mu \tag{22}$$

where a is the intercept and b the slope. Note that if the Gamma distribution was indeed an extremely accurate fit to probability density function of drop diameter at the ground p(D) then $\sigma = 0.5\mu\gamma$ which is not supported by the experimental evidences depicted in Figure 3. Eq.(22) suggests that only two parameters of the triplets (μ, σ, γ) are necessary to describe the probability density $p(D)$ of drop diameters at the ground. Thus one could define the rainfall phase space as a tridimensional space: e.g. the Cartesian product of the 3-tuple (N, μ, γ). However, longer data sets are necessary to effectively estimate the functions $a(\gamma)$ and $b(\gamma)$, and to explore the relationship $\sigma = a(\gamma) + b(\gamma)\mu$ in the range $\mu > 1$mm. Therefore, for the purpose of this manuscript, we conservatively consider $\mathbf{R}^4$ defined by the 4-tuple (N, μ, σ, γ) as the rainfall phase space.

4.2.3 Phase plots

To each 1 minute time interval of observation is associated the point of coordinates (N, μ, σ, γ) in the phase space. The entire data set occupies a volume in $\mathbf{R}^4$. To visualize this volume, we need to consider the six 2D projections: $\mu - \log_{10}(N)$, $\sigma - \log_{10}(N)$, $\gamma - \log_{10}(N)$, $\mu - \sigma$, $\mu - \gamma$, and $\sigma - \gamma$ (we use

$\log_{10}(N)$ instead of N for better visualization). Given a data set, we calculate the density of points in the phase space for all six 2D projections. Ten separate figures (one for each data set) would be necessary to illustrate the results. To obviate to this difficulty and give an idea to the reader of the differences/similarities between data sets we adopted the concept of average bounding perimeter. For all 2D projections we calculate the center of mass of the phase space points (each point has the same mass). With the center of mass as fixed point we span with an 10^o angle step the plane of the 2D projection. For each 10^o cone we calculate the average distance from the center of mass of the points within the cone. The connection with a continuous line of all average distance creates the average bounding perimeter which is a "measurement" of the volume of the phase space occupied by the database. The results are shown in Figure 4. The plots on the $\mu - \sigma$ projection plane show how the average bounding perimeters reflect the linear relationship between mean drop diameter and standard deviation of drop diameter depicted in Figure 3. The plots on the $\mu - \log_{10}(N)$ plane projection indicate that large value of counts ($\log_{10}(N) > 2.8$) are reached (on average) only for values of the mean drop diameter which are small (μ in the range 0.4-0.7 mm) or large ($\mu > 1mm$). The first case, many drops with small diameter, is a common feature of orographic precipitation as shown (e.g. [5], [6], [16], [1], and [4]). The BBY and CZC data sets are the same ones used in ([5], [6], and [16]) while BAO data set come from an instrument located at 1,577 m asml. The second case, many drops with possibly large diameters, is typical of strong convective events. the DRW data sets (rain of monsoonic origin) and the MIK (small island at the equator) and BKT (another equatorial site) are the data sets, among those considered, where the combination of large number of drops and large drop diameters occur more frequently on average. The average bounding perimeter on the $\mu - \gamma$ plane projection show that a decrease in value of the mean drop diameter is linked to a raise of the skewness value, although this may be in part an effect of the limitation of the instrument (Joss-Waldvogel impact disdrometer) which is not capable of detecting drop diameters smaller than 0.3 mm (reducing the contribution of left tails of $p(D)$ to the skewness). Due to the approximate linear relationship between μ and σ, results of projection on planes for which one of the axis is the standard deviation σ are similar to those for which axis is substituted by the mean μ. If we consider the result on the $\gamma - \log_{10}(N)$ plane projection, we see how these two variable are quite uncorrelated as the shape of the average bounding perimeters do not suggest any particular relation.

4.2.4 Rain rate and phase space parameters

The rainfall rate R aside from a multiplicative constant is the sum of three factors (Eq.21): $N\mu^3$, $3N\mu\sigma^2$, and $N\sigma^3\gamma$. Each factor accounts for a fraction $\alpha \in [-1, 1]$ (negative values are possible only for the factor $N\sigma^3\gamma$) of the rainfall rate. We calculate α for each factor and each 1 minute time interval of observation. Then we calculate the probability $F(\alpha)$ that the fraction does not exceed α. The results for each data base and each factor are shown in Figure 5. The $F(\alpha)$ curves are quite independent from the particular site. We see how the factor $N\mu^3$ contributes the most to the rainfall rate with a median contribution α_m ($F(\alpha_m) = 0.5$) which is in the range 0.7-0.75. The second largest contribution comes from the factor $3N\mu\sigma^2$, α_m in the range 0.2-0.25, while the factor $N\sigma^3\gamma$ accounts for the smallest contribution: $\alpha_m$in the range 0.025-0.05. These results can be explained noticing that at all sites and for all time interval of observation: 1) $\sigma < 1$, and 2) $\sigma < \mu$. So that $3\mu\sigma^2$ is almost always smaller than μ^3, and while γ can be larger (in absolute value) than μ, $\sigma^3\gamma$ is always smaller than μ^3 in virtue of 1) and 2).

5 Conclusions

When an objective measure of fit goodness is adopted, the Gamma distribution provides a poor fit to the drop size distribution sampled at short time scale (1 minute in our case) at all the ten sites considered. It is the opinion of the Authors that only an objective criterion of fit goodness (e.g. Kolmogorov-Smirnov) should guide the choice of a particular functional form for the concentration $\mathcal{N}(D)$ and/or the probability density function at the ground $p(D)$. For this reason we reject the Gamma distribution as a proper parametrization of the rainfall phenomenon. We propose an alternative parametrization based on the common statistical procedure of describing an unknown probability density function in term of its standardized central moments. We show that the 4-tuple of parameters (N, μ, σ, γ) is sufficient to describe the observed variability of disdrometer counts for all the ten sites considered, and refer to the Cartesian product of 4-tuple (N, μ, σ, γ) as the rainfall phase space. The volumes in the phase space relative to each data base (Figure 4) have some common features and some discrepancies which reflect different synoptic conditions and/or mechanisms of drop productions at play at the different site considered. However some results remarkably independent

from the site considered: 1) standardized central moments of order ≥ 4 have strong deterministic relationship with the third standardized moment: the skewness. 2) mean μ, standard deviation σ, and skewness γ are related to each other via Eq.(22) with values of the slope $b(\gamma)$ and intercept $a(\gamma)$ which are not compatible with a Gamma distribution functional dependence. Finally, bulk variables of the rainfall can be written as a function of 4-tuple (N, μ, σ, γ). e.g. Eq.(21) for the rainfall rate R. Bulk variables, such as the liquid water content W and the reflectivity Z are proportional to fractional moments (2.33 and 5.33 respectively) of the probability density function of drop diameter at the ground $p(D)$. Therefore, analytical expressions for the variables W ans Z in terms of the 4-tuple $(N, \mu, \sigma.\gamma)$ are necessarily approximations, on the contrary the Gamma distribution approximation leads to exact analytical expression. However, the main result of this paper is that any approximated expression, obtained via the proposed parametrization, is physically meaningful while any exact expression, obtained via the Gamma distribution parametrization, is not.

6 Acknowledgment

We wish to thank Dr.C.R.Williams and the National Oceanic and Atmospheric Administration (public availability of the data sets recorded at Bodega Bay, Eire, Cazadero, Darwin, and Kwajlein), the National Institute of Information and Communications Technology, Japan (Kashima data set), the Institute of Observational Research for Global Change together with the Japan Agency for Marine-Earth Science and Technology (Bukit Koto Tabang data set), Dr. Dan Brawn (Hassel data set), Dr. Martin Hagen of the Institut fuer Physik der Atmosphaere Deutsches Zentrum fuer Luft und Raumfahrt, Wessling, Germany (Macunaga data set), and the British Atmospheric Data Centre, Chilbolton data archive (Chilbolton data set).

Table 1: List of the sites from which Joss-Waldvogel disdrometer data are considered, with a three letters code for short referral and the Köppen-Geiger climate classification.

Site	**Code**	**Köppen-Geiger climate classification**
Eire, Colorado (USA)	BAO	Snow climate, fully humid with warm summer
Bodega Bay, California (USA)	BBY	Warm temperate climate with dry and warm summer
Bukit Koto Tabang, Indonesia	BKT	Equatorial rain forest, fully humid
Chilbolton, United Kingdom	CHB	Warm temperate climate, fully humid with cool summer and cold winter
Cazadero, California (USA)	CZC	Warm temperate climate with dry and warm summer
Darwin, Australia	DRW	Equatorial savannah with dry winter
Hassel, Germany	HSL	Warm temperate climate, fully humid with warm summer
Kashima, Japan	KSH	Warm temperate climate with dry and warm summer
Macunaga, Italy	MAC	Snow climate, fully humid with cool summer and cold winter
Kwajalein Atoll, RMI	MIK	Equatorial rainforest, fully humid

Table 2: List of the sites (short code referral) from which Joss-Waldvogel disdrometer data considered with latitude, longitude, altitude, number of 1 minute time interval in data set, and total number of drops in the data set. The symbol (*) indicates quantities calculated after data sets are processed according to the procedure described in Section 2.1.

Code	**Long.**	**Lat.**	**Alt. (m)**	**#min***	**#drop***
BAO	40.05N	105.00W	1,577	6,016	2,349,280
BBY	38.20N	123.00W	12	10,804	5,389,240
BKT	0.12S	100.19E	864	68,389	25,109,376
CHB	51.14N	1.43W	82	29,122	7,015,480
CZC	38.61N	123.22W	475	76,137	44,252,384
DRW	12.45S	130.83E	12	6,863	2,753,037
HSL	51.51N	7.1E	60	26,402	7,072,649
KSH	35.95S	140.65E	45	68,570	19,752,935
MAC	45.97N	7.96E	1,300	9,956	3,275,264
MIK	8.71N	167.73W	1	20,170	7,594,915

Table 3: Percentage of acceptance and rejection of the (lower truncated) Gamma distribution using the Kolmogorov-Smirnov goodness-of-fit test to each minute of the DRW data (6863) with a 5% level of significance. In columns 2 and 3 the critical value used is $1.3581/\sqrt{N}$ assuming that no parameters of the Gamma distribution are estimated, while in columns 4 and 5 the critical value is determined via Montecarlo simulations taking into account the fact that the parameters of Gamma are estimated via the MML from data. The results refer to entire DRW dataset (first row), and to subsets obtained considering as thresholds, the percentiles of the distribution of number of drops N per minute (second tttlast row). The thresholds are $N_{5\%} = 66$, $N_{25\%} = 104$, $N_{50\%} = 182$, $N_{75\%} = 455$, and $N_{95\%} = 1589$.

	ACP	RJC	ACP	RJC
ALL	71%	29%	**45%**	**55%**
$< N_{5\%}$	94%	6%	**78%**	**22%**
$\geq N_{5\%}$	70%	30%	**44%**	**56%**
$< N_{25\%}$	93%	7%	**71%**	**29%**
$\geq N_{25\%}$	64%	36%	**37%**	**63%**
$< N_{50\%}$	91%	9%	**65%**	**35%**
$\geq N_{50\%}$	52%	48%	**25%**	**75%**
$< N_{75\%}$	84%	16%	**56%**	**44%**
$\geq N_{75\%}$	34%	66%	**13%**	**87%**
$< N_{95\%}$	74%	26%	**47%**	**53%**
$\geq N_{95\%}$	24%	76%	**5%**	**95%**

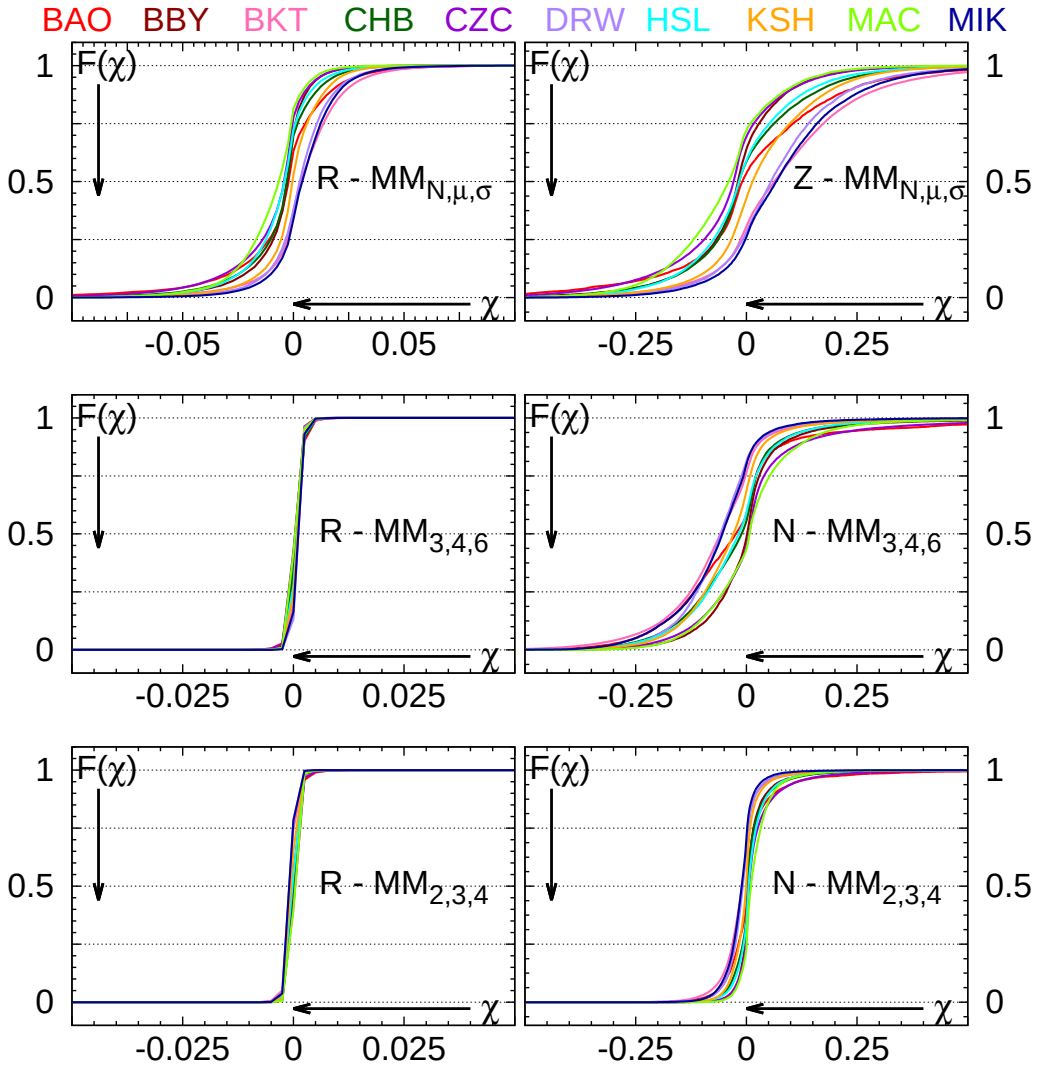

Figure 1: The observed frequency $F(\chi)$ of exceeding the relative error χ when calculating different bulk variables with the fitted Gamma distribution to the probability density function $p(D)$ of drop diameters at the ground. Variables R and Z for the $\mathrm{MM}_{N,\mu,\sigma}$ method (upper panels), variables R and N for the $\mathrm{MM}_{3,4,6}$method (middle panels), and $\mathrm{MM}_{2,3,4}$ method (bottom panels). Different colors indicate different data bases.

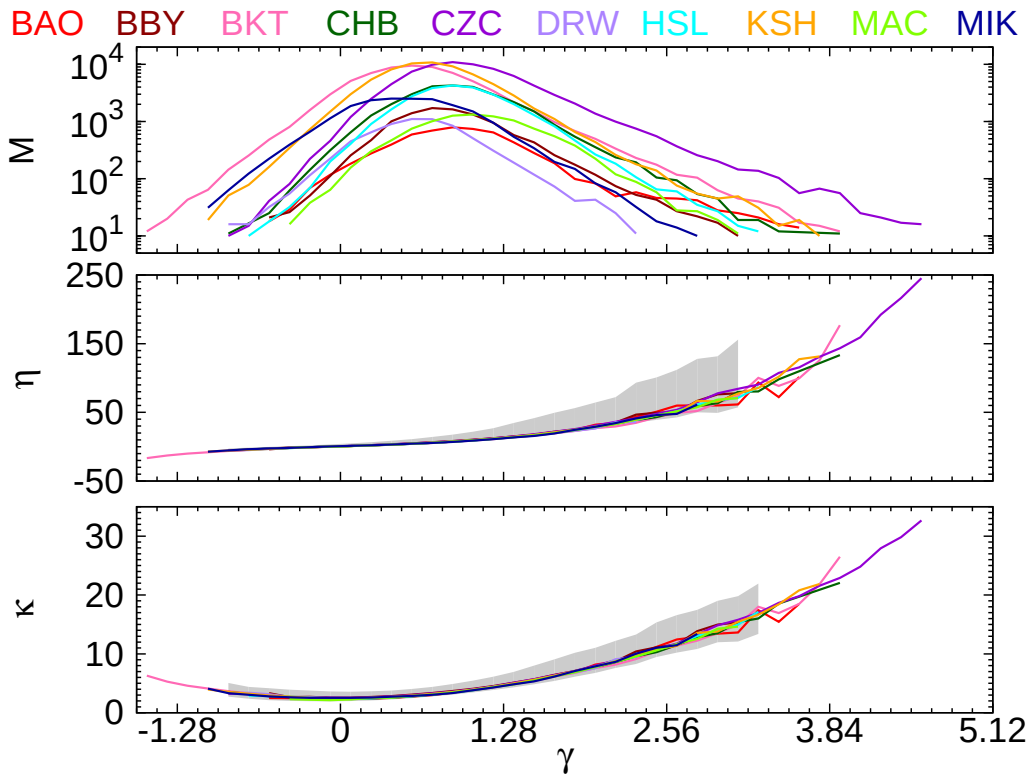

Figure 2: The number of samples M (upper panel), the kurtosis κ (middlepanel), and the fifth standardized central moment η (lower panel)as a function of the skewness γ. Solid lines indicate the number of samples and the median values for the parameters κ and η. Gray shadowed areas represent the 5%-, 95%-percentile range. Different colors indicate different data bases.

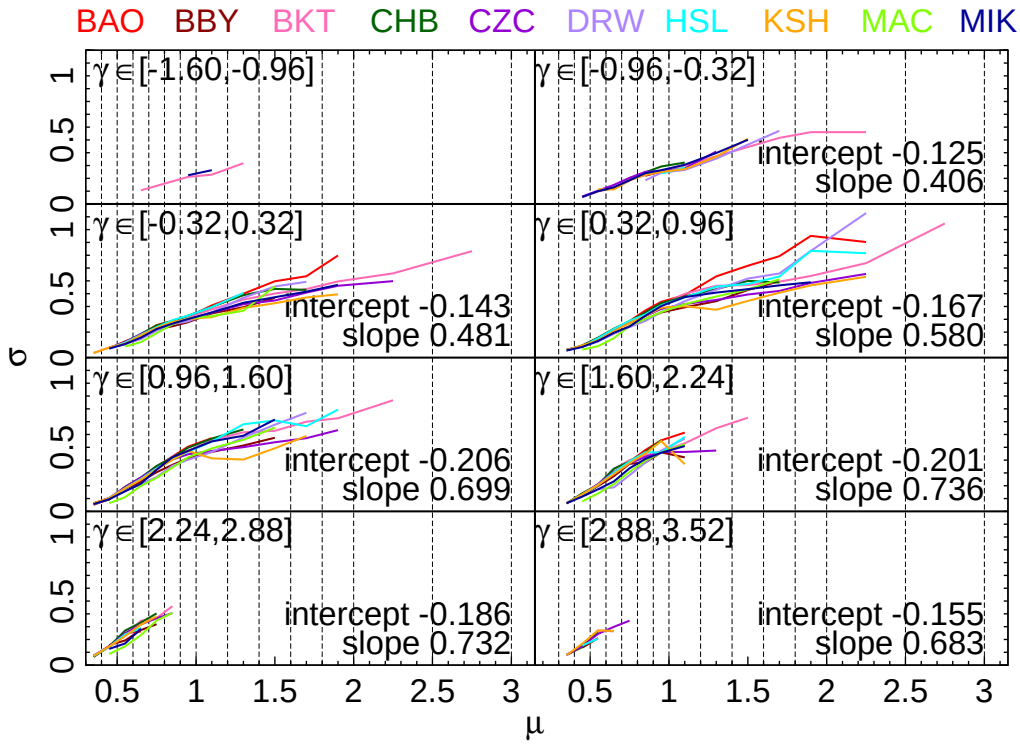

Figure 3: The median values of the standard deviation σ of drop diameters for different range value of the mean diameter μ (indicated by vertical dashed lines) and different range values of the skewness of drop diameter γ (shown in the top left corner of each panel). Also shown in the bottom right corner of each panel is the site average slope and intercept of the median standard deviation curves in the interval $\mu \in [0.3 : 1]$. Different colors indicate different databases.

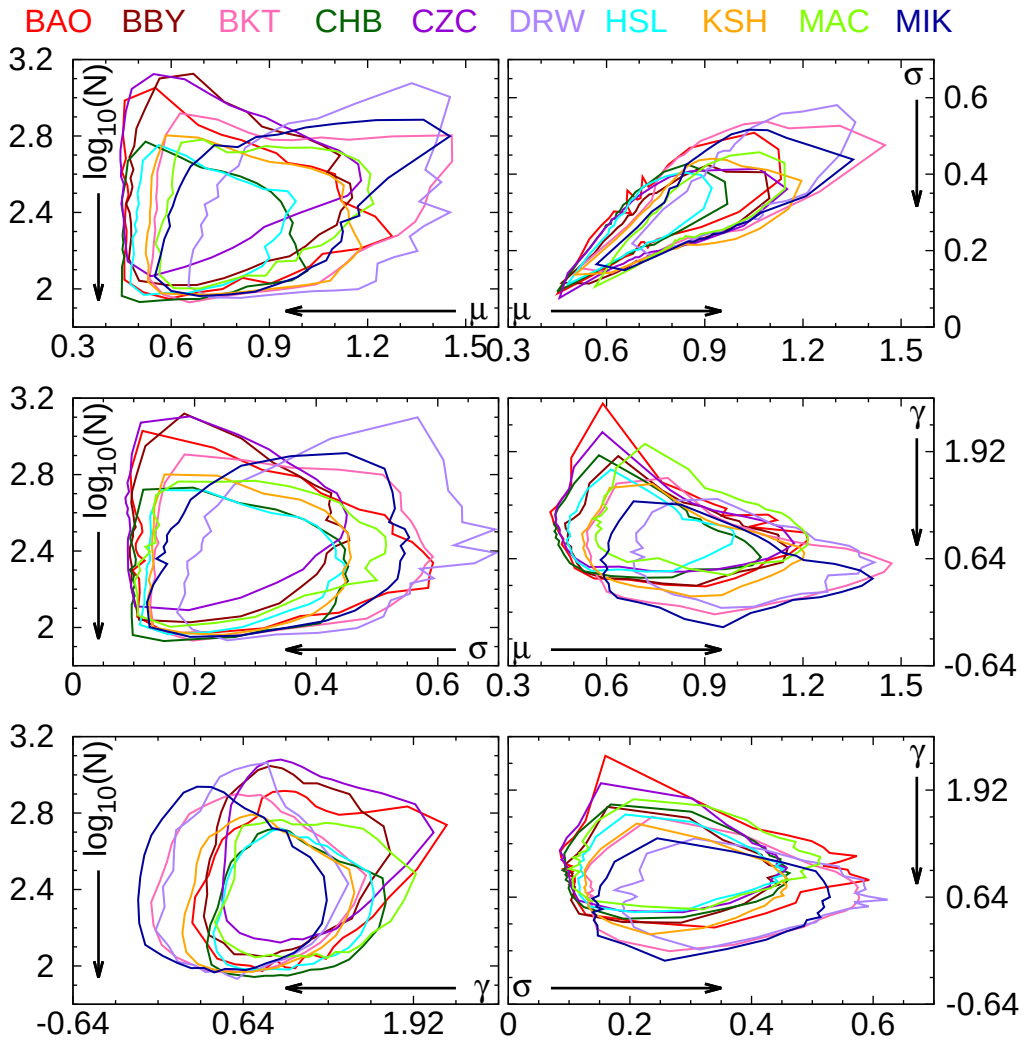

Figure 4: Phase space 2D projections plots of the average bounding perimeter lines for each site of observation. Different colors indicate different data bases.

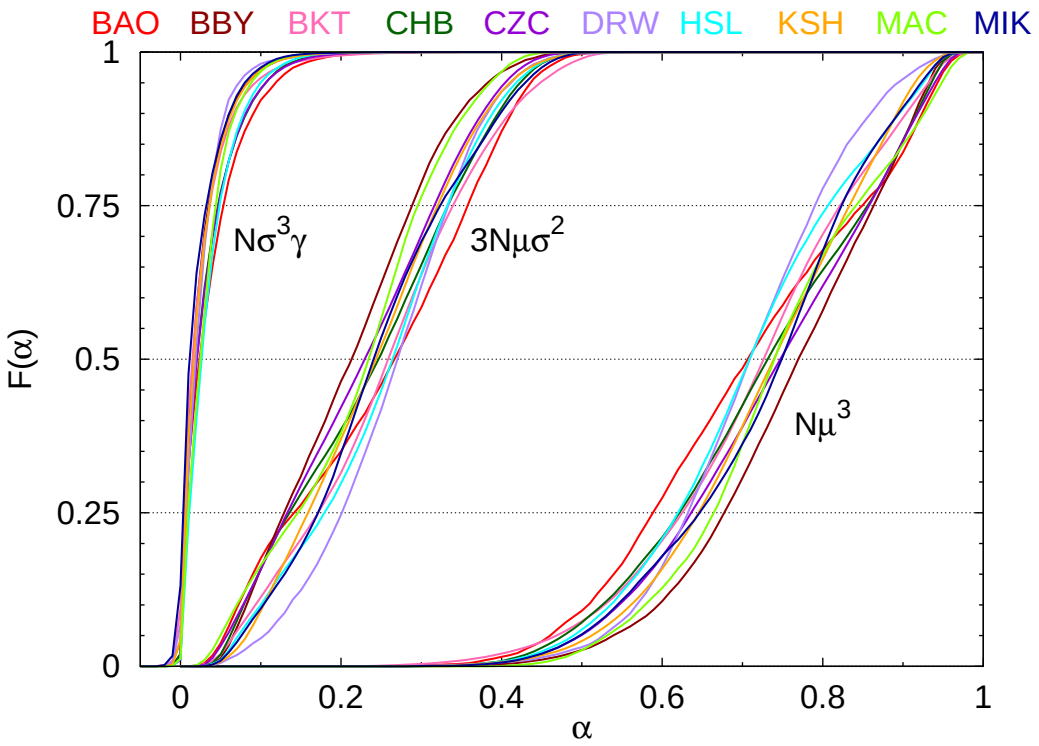

Figure 5: The probability $F(\alpha)$ of contributing a fraction α to the rainfall rate for the three factors $N\mu^3$, $3N\mu\sigma^2$, and $N\sigma^3\gamma$ of Eq.(21). Different colors indicate different data bases.

References

[1] Blanchard, D. C., 1953: Raindrop size-distribution in hawaiian rains. *J. of Meteorol.*, **10**, 457–473.

[2] Edwards, A., 1972: *Likelihood*. Cambridge University Press.

[3] Feingold, G. and Z. Levin, 1986: The lognormal fit to raindrop spectra from frontal convective clouds in israel. *Journal of Climate and Applied Meteorology*, **25 (10)**, 1346–1363.

[4] Fujiwara, M., 1967: Raindrop size distribution in warm rain as measured in hawaii. *Tellus XIX*, **3**, 392–402.

[5] Ignaccolo, M. and C. De Michele, 2012a: Skewness as measure of the invariance of instantaneous renormalized drop diameter distributions – part 1: Convective vs. stratiform precipitation. *Hydrology and Earth System Sciences*, **16 (2)**, 319–327.

[6] Ignaccolo, M. and C. De Michele, 2012b: Skewness as measure of the invariance of instantaneous renormalized drop diameter distributions – part 2: Orographic precipitation. *Hydrology and Earth System Sciences*, **16 (2)**, 329–343.

[7] Ignaccolo, M., C. De Michele, and S. Bianco, 2009: The droplike nature of rain and its invariant statistical properties. *Journal of Hydrometeorology*, **10 (1)**, 79–95.

[8] Johnson, R., D. Kliche, and P. Smith, 2009: Maximum likelihood estimation from a left-truncated distribution. *34th Conference on Radar Meteorology*, A. M. Soc., Ed., P2.2.

[9] Joss, J. and E. Gori, 1978: Shapes of raindrop size distributions. *J. Appl. Meteorol.*, **17**, 1054–1061.

[10] Keutelian, H., 1991: The kolmogorov-smirnov statistic when estimating parameters from data. CDF note 1285, Fermilab.

[11] Kliche, D., P. Smith, and R. Johnson, 2008: L-moment estimators as applied to gamma drop size distributions. *J. Appl. Meteorol. and Clim.*, **47**, 3117–3129.

[12] Kottegoda, N. and R. Rosso, 1997: *Statistics, Probability and Reliability for Civil and Environmental Engineers*. McGraw- Hill.

[13] Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the köppen-geiger climate classification updated. *Meteorol. Z.*, **15**, 259–263.

[14] Lavergnat, J. and P. Golé, 2006: A stochastic model of raindrop release: Application to the simulation of point rain observations. *J. Hydrol.*, **328**, 8–19.

[15] Marshall, J. and W. Palmer, 1948: The distribution of raindrops with size. *Journal of Meteorology*, **5**, 165–166.

[16] Martner, B. E., S. E. Yuter, B. W. Allen, S. Y. Matrosov, D. E. Kingsmill, and F. M. Ralph, 2008: Raindrop size distributions and rain characteristics in california coastal rainfall for periods with and without a radar bright band. *J. Hydrometeor.*, **9**, 408–425.

[17] Owolawi, P., 2011: Raindrop size distribution model for the prediction of rain attenuation in durban. *Progress In Electromagnetics Research Symposium Proceedings*, 1068–1075.

[18] Reed, M. and B. Simon, 1975: *Fourier Analysis, Self-Adjointness, Methods of modern mathematical physics*. Academic Press.

[19] Rosenfeld, D. and C. Ulbrich, 2003: Cloud microphysical properties, processes, and rainfall estimation opportunities. *Meteorological Monographs*, **30 (52)**, 237–258.

[20] Smith, J., E. Hui, M. Steiner, M. L. Baeck, W. Krajewski, and A. Ntelekos, 2009: Variability of rainfall rate and raindrop size distributions in heavy rain. *Water Resour. Res.*, **45**, W04 430.

[21] Smith, J. A., 1993: Marked point process models of raindrop-size distributions. *J. Appl. Meteor.*, **32**, 284–296.

[22] Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of 'normalized' distribution to descrbie raindrop spectra: A tool for cloud physics and cloud remote sensing. *J. Appl. Meteor.*, **40 (6)**, 1118–1140.

[23] Tokai, A. and D. A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.*, **35**, 355–371.

[24] Uhlenbeck, G. E. and L. S. Ornstein, 1930: On the theory of the brownian motion. *Physical Review*, **36 (5)**, 823–841.

[25] Ulbrich, C., 1983: Natural variations in the analytical form of the raindrop size distribution. *J. Climate Appl. Meteor.*, **22**, 1764– 1775.

[26] Villermaux, E. and B. Bossa, 2009: Single-drop fragmentation determines size distribution of raindrops. *Nature Physics*, **5**, 697–702.