

## A critical evaluation of the adequacy of the gamma model for representing raindrop size distributions

Gatidis, Christos; Schleiss, Marc; Unal, Christine; Russchenberg, Herman

**DOI**

[10.1175/JTECH-D-19-0106.1](https://doi.org/10.1175/JTECH-D-19-0106.1)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Journal of Atmospheric and Oceanic Technology

**Citation (APA)**

Gatidis, C., Schleiss, M., Unal, C., & Russchenberg, H. (2020). A critical evaluation of the adequacy of the gamma model for representing raindrop size distributions. *Journal of Atmospheric and Oceanic Technology*, 37(10), 1765-1779. <https://doi.org/10.1175/JTECH-D-19-0106.1>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

## A Critical Evaluation of the Adequacy of the Gamma Model for Representing Raindrop Size Distributions

CHRISTOS GATIDIS, MARC SCHLEISS, CHRISTINE UNAL, AND HERMAN RUSSCHENBERG

*Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, Netherlands*

(Manuscript received 22 June 2019, in final form 2 August 2020)

**ABSTRACT:** The adequacy of the gamma model to describe the variability of raindrop size distributions (DSD) is studied using observations from an optical disdrometer. Model adequacy is checked using a combination of Kolmogorov–Smirnov goodness-of-fit test and Kullback–Leibler divergence and the sensitivity of the results to the sampling resolution is investigated. A new adaptive DSD sampling technique capable of determining the highest possible temporal sampling resolution at which the gamma model provides an adequate representation of sampled DSDs is proposed. The results show that most DSDs at 30 s are not strictly distributed according to a gamma model, while at the same time they are not far away from it either. According to the adaptive DSD sampling algorithm, the gamma model proves to be an adequate choice for the majority (85.81%) of the DSD spectra at resolutions up to 300 s. At the same time, it also reveals a considerable number of DSD spectra (5.55%) that do not follow a gamma distribution at any resolution (up to 1800 s). These are attributed to transitional periods during which the DSD is not stationary and exhibits a bimodal shape that cannot be modeled by a gamma distribution. The proposed resampling procedure is capable of automatically identifying and flagging these periods, providing new valuable quality control mechanisms for DSD retrievals in disdrometers and weather radars.

**KEYWORDS:** Rainfall; Drop size distribution; Sampling

### 1. Introduction


Continuous and reliable precipitation monitoring is of fundamental importance for understanding the water cycle. Every year, extreme precipitation events cause floods and trigger landslides, which cost many human lives and billions of dollars (Ralph et al. 2014). However, obtaining accurate precipitation measurements can be extremely challenging due to the high underlying variability of the meteorological phenomenon in space and time (Jameson and Kostinski 2001; Uijlenhoet et al. 2003).

One fundamental quantity needed to understand rainfall variability is the raindrop size distribution (DSD). The DSD is considered to be the key source of uncertainty in quantitative precipitation estimations (QPE), affecting rain-rate estimates from ground based radars and satellites. Several studies have shown that QPE can be significantly improved using accurate DSD observations (Rose and Chandrasekar 2006) but is difficult in practice as natural DSDs rapidly vary in space and time and exhibit a wide range of shapes. For radar and satellite related applications, where a limited amount of information is available, it is often necessary to parameterize the DSD in the form of a simple distribution.

Several mathematical models have been proposed to approximate naturally occurring DSDs (Marshall and Palmer 1948; Ulbrich 1983; Zhang et al. 2001; Testud et al. 2001; Bringi et al. 2003). The most popular and widely accepted of them in the remote sensing community is the gamma distribution.

However, the gamma distribution is not a perfect model and several studies have questioned its adequacy (Kliche et al. 2008; Ekerete et al. 2015; Cugerone and De Michele 2015). Its acceptance mainly comes from the fact that it is relatively versatile yet simple enough to be useful in practice. It is more flexible than the exponential (Marshall and Palmer 1948) and provides a “reasonably good fit” to measured DSDs. In addition to the conventional distributions like gamma, more complex models have also been proposed in the literature (Ignaccolo and De Michele 2014; Ekerete et al. 2015; Cugerone and De Michele 2015; Thurai and Bringi 2018). Although they are better at representing real DSDs, they are more difficult to use in practice due to their large number of parameters that cannot be retrieved using remote sensing measurements.

Despite its being the most widely used model, only a few studies have focused on precisely quantifying the adequacy of the gamma distribution. Johnson et al. (2015) compared the performance of four conventional unimodal and skewed-to-the-right distributions that may be considered as potential DSD models. They highlighted that the gamma model provided the best fit, followed by the lognormal, beta and, finally, the Weibull. Adirosi et al. (2016) fitted three distributions—lognormal, gamma, and Weibull—and compared their goodness of fit using the Kolmogorov–Smirnov test (K-S test). They showed that the gamma distribution has the lowest rejection rate, while Weibull is the most frequently rejected. Ekerete et al. (2015) used the chi-square goodness-of-fit test for testing several candidate models against the observations and concluded that DSDs are somewhere between the bimodal and the gamma shape, suggesting that gamma or lognormal distributions are not fully adequate. Their recommendation is to use a Gaussian mixture model with three centers. Similarly, Cugerone and De Michele (2015) pointed

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Christos Gatidis, c.gatidis@tudelft.nl

DOI: 10.1175/JTECH-D-19-0106.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

out that the gamma and lognormal model are not accurate enough based on K-S test and a skewness–kurtosis diagram and suggested to use a four-parameter distribution (i.e., Johnson SB) instead. That diagram is one of the various graphical methods for determining visually whether sample data conform to a reference distribution. Among them the most commonly used graphical tools are the quantile-to-quantile plots or Q–Q plots (Yakubu et al. 2014; Watanabe and Ingram 2016) and the density plots (Ekerete et al. 2015; Adirosi et al. 2016) as well.

None of these studies focused on sampling resolution as a major factor nor provided clear guidelines for how it should be taken into account when evaluating the gamma model. All of them used a similar resolution (1-min DSDs), ignoring questions like “Does the gamma model perform better at lower/higher resolutions than 60 seconds?” or “When should we not use a gamma model?” These are very relevant questions when we take into consideration that DSD measurements at higher temporal resolutions are affected by larger sampling uncertainties. Although it was not the main scope of their study, Adirosi et al. (2015) suggested that the adequacy of the gamma model is likely to decrease with higher temporal resolutions.

In this paper, we take a closer look at the adequacy of the gamma model for representing DSDs. Our analysis starts by fitting gamma distributions on a DSD dataset collected by an optical disdrometer over a 2-month period. The adequacy of the fit is assessed based on a combination of Kolmogorov–Smirnov goodness-of-fit test and Kullback–Leibler divergence. The novelty of the study lies in its focus on the influence of the sampling resolution and how it affects the adequacy of the gamma model. A new adaptive DSD sampling technique capable of determining the highest possible temporal sampling resolution at which the gamma model provides an adequate fit is proposed.

The work is organized as follows. In section 2, we introduce the data used in our study, and in section 3 we present the followed methodology. The adequacy of the gamma model and its sensitivity to the sampling resolution is presented in section 4. Finally, the conclusions are provided in section 5.

## 2. Data

The DSD data used in this study were collected by a Particle Size and Velocity (Parsivel<sup>2</sup>) optical disdrometer located at the Cabauw Experimental Site for Atmospheric Research (CESAR) observatory during the Analysis of the Composition of Clouds with Extended Polarization Techniques (ACCEPT) campaign in October and November 2014, in a collaboration between Leibniz Institute for Tropospheric Research (TROPOS), Germany, and Delft University of Technology. Cabauw is located in the western part of the Netherlands, in a polder area, 0.7 m below mean sea level.

The measurement principle and performance of the Parsivel disdrometer have been extensively described in previous studies (Löffler-Mang and Joss 2000; Tokay et al. 2014); it uses a horizontal laser beam with an approximately 54-cm<sup>2</sup> measuring, sampling surface. When a raindrop passes through the laser beam, the attenuation in the received voltage and the time for the particle to leave the beam can be used to estimate

the equivolume spherical raindrop diameter and the terminal fall velocity of the raindrop.

Diameter and velocity are divided in 32 nonuniformly spaced classes ranging from 0 to 25 mm and 0 to 21 m s<sup>−1</sup> respectively. All the drops in a given class are assigned to the center of the bin. Thus, the raw output data for a sampling interval are a 32 × 32 matrix of detected number of drops for each diameter and velocity class from which the volumetric size distribution of drops (m<sup>−3</sup> mm<sup>−1</sup>) can be estimated (Raupach and Berne 2015). The first two diameter classes (0.062 and 0.187 mm) in the Parsivel are always empty due to the low signal-to-noise ratio. Therefore, the minimum detectable drop diameter is approximately 0.25 mm. From the DSD, integrated quantities can be inferred, such as rainfall rate ( $R$ ) expressed in mm h<sup>−1</sup>, which is the amount of rain that falls over a given interval of time and radar equivalent reflectivity factor ( $Z$ ) expressed in dBZ which is related to the back-scattered radar signal of hydrometeors. For this study, volumetric DSD values at several temporal sampling resolutions were used, the highest being 30 s.

As in other related DSD studies (Thurai and Bringi 2018; Jaffrain and Berne 2011; Adirosi et al. 2014) a selection criterion is applied to the whole DSD dataset before the analysis. Our selection is based on two main requirements:

- Only liquid precipitation is considered. This means that only the first 22 diameter classes of the Parsivel are used, starting from the third (0.25–0.375 mm) up to the twenty-second (6–7 mm), the latter corresponding to the biggest physically possible raindrops. DSDs with measurements between the twenty-third and thirty-second class (solid or mixed precipitation) were discarded. Further, the classification of precipitation particles provided by Parsivel was used, and only the following precipitation types were accepted: drizzle, drizzle with rain and rain. All DSDs attributed to other precipitation types were discarded.
- A threshold on the minimum number of size classes was set. All DSDs composed of fewer than three different size classes were discarded. Moreover, all DSDs for which the rain rate estimated by the Parsivel is zero were discarded. This is necessary in order to remove suspicious observations such as noise from insects or other objects crossing the beam.

The application of the selection procedure described above to the whole dataset (24 289 DSD spectra) resulted in 12 329 30-s DSD spectra. From the 11 960 DSD that were excluded, more than 80% were discarded due to the spurious signals (noise) with only one diameter class and around 1.5% because of solid or mixed precipitation. In Table 1 we present the different percentiles of rain rate, reflectivity factor, mean diameter and number concentration corresponding to the whole dataset. These values indicate that the dataset mostly contains light rain, with 90% of the time steps corresponding to rain rates of less than 2 mm hr<sup>−1</sup>, which in terms of reflectivity factor is less than 30 dBZ and in terms of mean drop sizes less than 1.33 mm. This can be explained if we take into account the fact that the ACCEPT campaign took place in October–November in the Netherlands, during the cold season. Hence most of the rain events are frontal in nature and convective

TABLE 1. Overview of the DSD dataset from Parsivel<sup>2</sup> optical disdrometer during the ACCEPT campaign. Sampling resolution, available number of DSD samples, accepted number of samples after applying the filtering criteria, mean, 50%, 90%, and 99.9% quantile and maximum value of rain intensity, reflectivity factor, mass-weighted mean diameter, and number concentration.

Sampling resolution	No. of DSD spectra			No. of samples after selection
30 s	24 289			12 329
	Rain intensity ( $R$ ) ( $\text{mm h}^{-1}$ )	Reflectivity factor ( $Z$ ) (dBZ)	Mass-weighted mean diameter ( $D_m$ ) (mm)	Number concentration ( $N_T$ ) ( $\text{m}^{-3}$ )
Mean	0.76	16.17	0.89	401
50.0% (median)	0.33	16.20	0.82	225
90.0%	2.02	29.36	1.33	700
99.9%	11.20	42.08	3.12	8263
Max	26.31	47.82	4.92	11 193

events are rare. However, there are also some larger values with rain rates between 11.2 and 26.31  $\text{mm h}^{-1}$ , number concentration between 8263 and 11 193 drops per cubic meter and mean diameter between 3.12 and 4.92 mm. Note that the maximum mean diameter (4.92 mm) is suspicious and probably corresponds to solid precipitation incorrectly classified by the disdrometer highlighting the limitations of the Parsivel in terms of detection of small/big droplets (Raupach and Berne 2015; Tokay et al. 2014).

For illustration and study purposes, one particularly interesting event during the campaign was chosen (12 October 2014). This event was selected because it exhibits a large natural variability in DSD. As can be seen from Fig. 1, the event started at 1900 UTC and lasted almost 4 h. According to the measured DSDs in Fig. 2, the event can be divided into three separate parts, each of them with its own characteristics. The first part between 1900 and 2000 UTC is characterized by a relatively stationary DSD, with no remarkable changes in the number concentration (between 200 and 300 raindrops per cubic meter) and a mean diameter mainly between 1 and 1.5 mm. During the second part between 2040 and 2120 UTC, the mean drop size increases and the DSD becomes more variable, with rain rates between 1.5 and 7.5  $\text{mm h}^{-1}$ . The last part between 2158 and 2237 UTC exhibits a more stable DSD with a large number of small drops below 1 mm (drizzle conditions). It should be noted that even though the number concentration has a peak around 2211 UTC ( $N_{T,\text{max}} = 2698 \text{ m}^{-3}$ ), the equivalent reflectivity factor and the precipitation intensity peak before 2100 UTC, highlighting the larger sensitivity of both variables to the drop sizes rather than concentration.

### 3. Methodology

Disdrometers provide detailed information about the DSD and its variation over time. However, when we need to derive DSDs using radar or satellite data, because of the limited set of available observables, it is necessary to summarize this information in the form of a mathematical model. Unfortunately, it is impossible to find simple models that perfectly capture the complex, natural properties of observed DSDs. One of the most common DSD models used in practice is the gamma model by Ulbrich (1983):

$$N(D) = N_0 D^\mu e^{-\Lambda D}, \quad (1)$$

where  $N_0$  is the intercept parameter ( $\text{m}^{-3} \text{mm}^{-1-\mu}$ );  $\mu$  is the shape parameter (unitless) and  $\Lambda$  is the slope parameter ( $\text{mm}^{-1}$ ). To overcome the dependence of  $N_0$  to  $\mu$ , the model is often reformulated and normalized (Testud et al. 2001; Bringi et al. 2003):

$$N_{\text{model}}(D) = N_w f(\mu) \left( \frac{D}{D_m} \right)^\mu e^{-(4+\mu)D/D_m}, \quad (2)$$

with

$$f(\mu) = \frac{6}{4^4} \frac{(\mu+4)^{(\mu+4)}}{\Gamma(\mu+4)}, \quad (3)$$

$$\Lambda = \frac{4+\mu}{D_m}, \quad (4)$$

where  $D_m$  (mm) is the mass-weighted mean diameter and  $N_w$  the generalized intercept parameter ( $\text{m}^{-3} \text{mm}^{-1}$ ) whose unit does not depend on  $\mu$ . Various methods have been proposed to estimate gamma DSD parameters ( $N_w$ ,  $\mu$ ,  $D_m$ ) from disdrometer observations. In this study we focus on the two most common: the method of moments (MoM) and maximum likelihood estimation (MLE).

#### a. MoM

The MoM is based on the fact that the unknown parameters of the fitted DSDs can be expressed as a combination of different weighted moments of the DSD (Ulbrich 1983). Different versions of MoM have been proposed depending on the application: the method of truncated moments (Ulbrich 1985), the L-moment method (Johnson et al. 2011; Kliche et al. 2008) and the  $\mu$ -search method applied to normalized spectra (Thurai et al. 2014). Here we focus on the latter. For each observed DSD provided by the Parsivel disdrometer,  $D_m$  and  $N_w$  are calculated:

$$D_m = \frac{\sum_{i=3}^{22} N(D_i) D_i^4 dD_i}{\sum_{i=3}^{22} N(D_i) D_i^3 dD_i}, \quad (5)$$

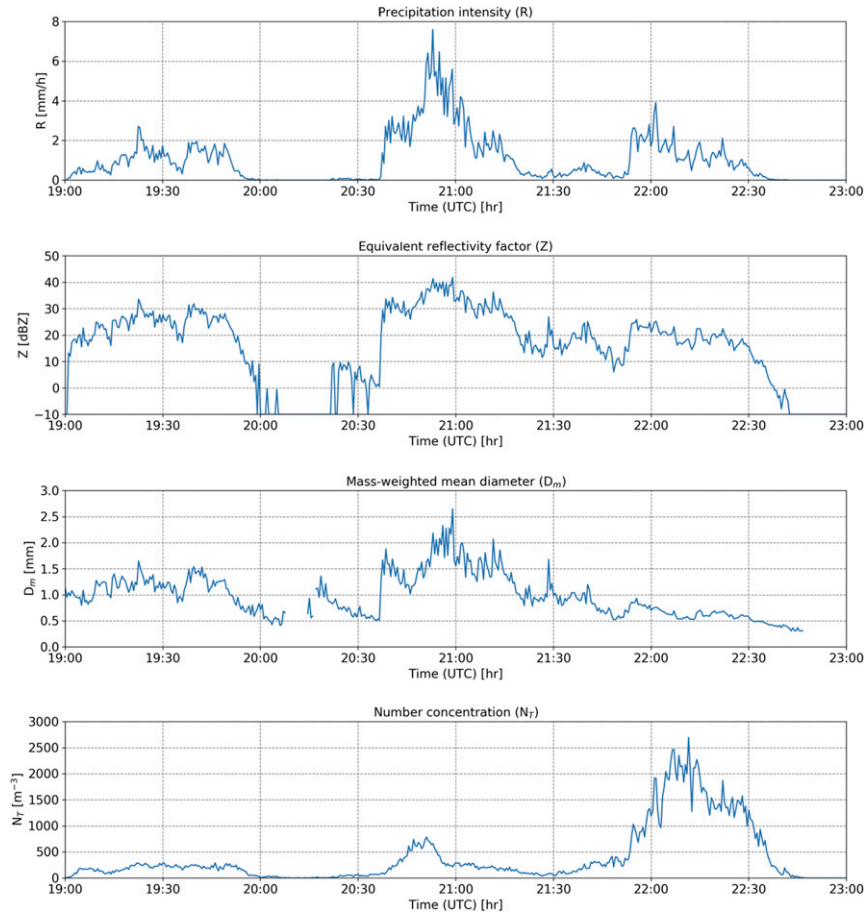


FIG. 1. Time series of (top to bottom) precipitation intensity ( $\text{mm h}^{-1}$ ), equivalent reflectivity factor (dBZ), mass-weighted mean diameter (mm), and number concentration ( $\text{m}^{-3}$ ) from Parsivel disdrometer data on 12 Oct 2014.

$$\text{LWC} = \frac{\pi \rho_w}{6} \sum_{i=3}^{22} N(D_i) D_i^3 dD_i, \quad (6)$$

$$N_w = \frac{4^4}{\pi \rho_w} \left( \frac{\text{LWC}}{D_m^4} \right), \quad (7)$$

where LWC denotes the liquid water content ( $\text{g m}^{-3}$ ),  $\rho_w$  is the density of water ( $10^{-3} \text{ gmm}^{-3}$ ),  $D_i$  is the center of the  $i$ th diameter class,  $N(D_i)$  is the measurement by Parsivel (volumetric size distribution), and  $dD_i$  is the width of the  $i$ th diameter class. We compute  $D_m$  as the ratio of the fourth to the third moment of the DSD. Using the calculated  $N_w$  and  $D_m$  values, the shape parameter  $\mu$  can be computed. The optimal shape parameter  $\mu$  is estimated for every individual time step, by determining the value of  $\mu \in [-3, 15]$  that minimizes the following cost function (Thurai et al. 2014):

$$\text{CF} = \sum_{i=3}^{22} |\log_{10}[N(D_i)] - \log_{10}[N_{\text{model}}(D_i|\mu)]|. \quad (8)$$

The applied brute force search tests all possible values of  $\mu$  between  $-3$  and  $15$  with steps of  $0.01$  and selects the one that

minimizes Eq. (8). Note that due to the large number of zeros in measured DSD spectra [ $N(D_i) = 0$ ] and the numerical issues related to  $\log_{10}(0)$  in Eq. (8), instead of the logarithms, square roots were used in this study.

#### b. MLE

The second method used to estimate the parameters of the gamma DSD model is the MLE. MLE is a statistical technique for evaluating how likely it is to observe a specific output under the assumption of a given set of model parameters. Several studies have applied MLE method for DSD retrievals (Adirosi et al. 2016; Thurai et al. 2014; Schleiss et al. 2009). The goal of MLE is to find a pair of  $\mu$  and  $\Lambda$  values ( $\mu^*$ ,  $\Lambda^*$ ) that maximizes the log-likelihood function:

$$\sum_{i=1}^n \ln[f(y_i; \mu^*, \Lambda^*)] = \max, \quad (9)$$

where  $y_1, y_2, \dots, y_n$  are the observations,  $f(y_i)$  denotes the density, and  $n$  is the total number of observations. In case of a gamma distribution  $f(y)$  is given by



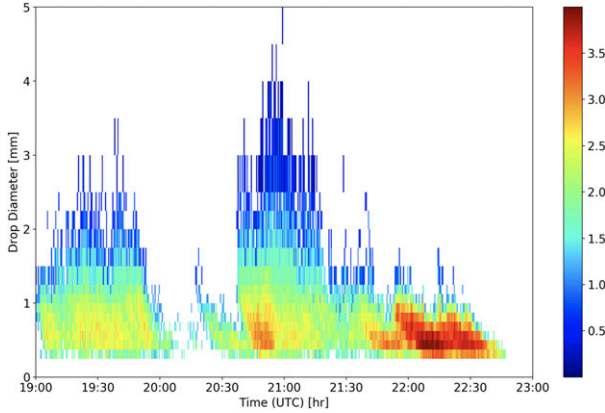


FIG. 2. Volumetric drop size distributions  $N(D)$  ( $\text{m}^{-3} \text{mm}^{-1}$ ) in logarithmic scale as a function of time for the study case on 12 Oct 2014.

$$f(y; \mu, \Lambda) = \frac{\Lambda^{\mu+1}}{\Gamma(\mu+1)} y^{\mu} e^{-\Lambda y}. \quad (10)$$

Note that because the Parsivel disdrometer outputs the number of drops per cubic meter and per diameter class  $i$  [ $N(D_i)$ ] and not the total number of observations,  $n$  is given by

$$n = \sum_{i=3}^{22} N(D_i). \quad (11)$$

To take the binning into account, Eq. (9) is rewritten in terms of the Parsivel observations:

$$\sum_{i=3}^{22} N(D_i) \ln[f(D_i; \mu^*, \Lambda^*)] = \max. \quad (12)$$

The MLE is the pair of parameters  $(\mu^*, \Lambda^*)$  that maximizes Eq. (12). This is determined through numerical optimization, for example by using a steepest gradient method. As initial values  $(\mu_0, \Lambda_0)$  for the optimization, the results from the MoM and Eq. (4) can be used.

One important limitation of the Parsivel data is the effect of censoring at the lower and higher drop end. Therefore, some studies have suggested to rescale the density function in the likelihood function over the range of observable diameter classes (Johnson et al. 2014). In this case the density in Eq. (10) becomes

$$f(D_i; \mu, \Lambda) = \frac{\int_{D_{i,\min}}^{D_{i,\max}} \Lambda^{\mu+1} y^{\mu} e^{-\Lambda y} dy}{\frac{\gamma(\mu+1, \Lambda D_{22,\max})}{\Gamma(\mu+1)} - \frac{\gamma(\mu+1, \Lambda D_{3,\min})}{\Gamma(\mu+1)}}, \quad (13)$$

where  $\gamma$  is the incomplete gamma function,  $D_{i,\max}$  and  $D_{i,\min}$  are the upper and lower boundary of the  $i$ th diameter bin (with  $i = 3, 4, \dots, 22$ ), respectively, with  $D_{22,\max} = 7.0$  mm and  $D_{3,\min} = 0.25$  mm. In this study, both approaches were used whereas it should be pointed out that the first (without rescaling) is significantly faster than the second. The MLE for truncated and binned data will be discussed further in appendix A.

### c. K-S test

As mentioned before, this study primarily focuses on the adequacy of the gamma model. This will be examined using two statistical tools. The first is the K-S test. The K-S test is a nonparametric test that quantifies the difference  $D_n$  between the empirical distribution function (ECDF) of the sample ( $F_n$ ) and the cumulative distribution function (CDF) of a reference distribution ( $F$ ):

$$D_n = \sup_x |F_n(x) - F(x)|, \quad (14)$$

where  $\sup$  is the supremum of the set of distances. From  $D_n$ , a  $p$  value can be calculated which is used to reject or accept the null hypothesis ( $H_0$ ) that the DSD sample comes from the theoretical gamma distribution (Cugerone and De Michele 2015; Adirosi et al. 2016). When the  $p$  value is smaller than the significance level ( $\alpha = 0.05$ ) the gamma model is rejected. To apply the K-S test, the function “scipy.stats.kstest” from the SciPy Python library was used.

At this point it is worth mentioning that because the K-S test can be applied only to continuous distributions, it is not directly applicable to a discrete dataset like the binned DSD observations provided by a Parsivel disdrometer. To overcome this issue, randomization of the drop sizes based on a uniform random distribution of the raindrops over each class was used (Ignaccolo and De Michele 2014; Chambers et al. 1983). The drops can also be redistributed according to the density function of a gamma distribution. But our analyses showed that this is much slower and does not significantly change the results.

Also, it should be noted that because the K-S test is applied to estimated model parameters, the mathematically correct but computationally expensive way to determine the  $p$  values is to use Monte Carlo simulations (Adirosi et al. 2016; Laio 2004; Ignaccolo and De Michele 2014). Again, both approaches were considered. For the application of the K-S test using Monte Carlo simulations (“exact” K-S test), the reader is referred to appendix B.

### d. Kullback–Leibler divergence

The second tool used to evaluate the adequacy of the gamma model is the Kullback–Leibler divergence ( $D_{\text{KL}}$ ) also known as relative entropy (Kullback and Leibler 1951):

$$D_{\text{KL}}(P \parallel Q) = \sum_{i=3}^{22} P(D_i) \ln \left[ \frac{P(D_i)}{Q(D_i)} \right], \quad (15)$$

$$P(D_i) = \frac{N(D_i)}{\sum_{i=3}^{22} N(D_i)}, \quad (16)$$

where  $P(D_i)$  denotes the probability distribution function (PDF) of the observations from the disdrometer and  $Q$  is the PDF of the reference distribution [gamma model, Eq. (10)]. The  $D_{\text{KL}}$  measures how much work needs to be made in order to transform one distribution into the other. Values close to 0 mean that  $P$  and  $Q$  are very similar, while a value of 1 indicates that they are completely different. In contrast to the K-S test output which is a “yes” or “no,” the  $D_{\text{KL}}$  provides a more

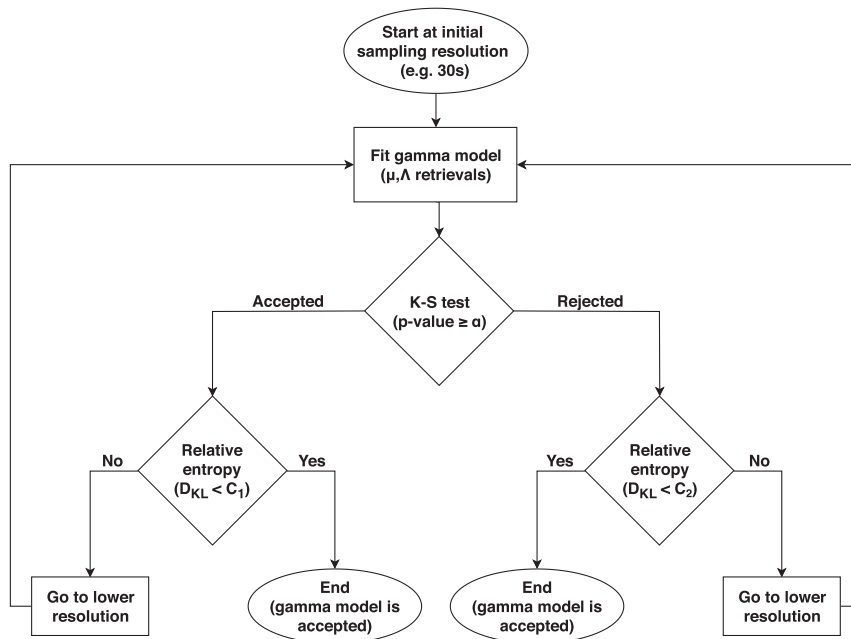


FIG. 3. Flowchart of the adaptive sampling algorithm for determining the highest possible resolution at which the gamma hypothesis is adequate.

nuanced answer in terms of how close or far away the modeled DSD is from our initial gamma DSD assumption. By combining K-S test and  $D_{KL}$ , a more detailed assessment of the adequacy of the gamma model can be made. Similarly to the K-S test,  $D_{KL}$  is calculated using the Python function “`scipy.stats.entropy`” from the SciPy library.

#### e. Adaptive sampling

Natural variability in rain means that the DSD is seldomly constant over time. However, for everything that follows, we assume local statistical stationarity in the DSD, i.e., that the unconditional joint probability function of raindrop sizes and number concentrations are invariant over small time intervals and spatial volumes. This is common practice and necessary for making inference, although questionable from an observational point of view (see, e.g., Jameson and Kostinski 2001; Ignaccolo et al. 2009; Schleiss et al. 2014; Gires et al. 2015). In fact, nonstationarity in rainfall could be one of the reasons why the gamma model is not equally good at representing DSDs across different aggregation time scales.

In addition to that, measurements can also be affected by significant uncertainties due to the limited sampling area of the Parsivel. As a result, the adequacy of the gamma model will change depending on the considered sampling resolution. To reveal this sensitivity, the original measurements at 30 s were resampled to lower temporal resolutions. Resampling was done by averaging the 30-s DSDs (i.e., summing consecutive DSD spectra and dividing by the number of measurements) backward in time, using overlapping 30-s windows.

An adaptive sampling technique is used to determine the highest possible temporal sampling resolution at which the gamma model provides an adequate fit. The adequacy is

assessed by using a combination of K-S test and Kullback–Leibler divergence ( $D_{KL}$ ). As can be seen schematically (Fig. 3), the algorithm is an iterative procedure which starts at the highest possible sampling resolution (e.g., 30 s). At this resolution, the gamma model is fitted and its parameters ( $\mu$ ,  $\Lambda$ ) are retrieved. The K-S test is applied to the fitted model and  $D_{KL}$  is calculated. Based on the acceptance or the rejection of the K-S test, the right or left branch of the decision tree is followed. In both cases, an additional test on the  $D_{KL}$  with thresholds  $C_1$  and  $C_2$  determines whether the algorithm ends or not. If  $D_{KL}$  is greater than  $C_1$  and the K-S test is accepted or if  $D_{KL}$  is greater than  $C_2$  and the K-S test rejects, the gamma model is rejected at this specific resolution and this procedure continues to a lower sampling resolution. The two other cases result in the gamma model being accepted and the algorithm moves on to the next DSD observation. The two thresholds  $C_1$  and  $C_2$  characterize how tolerant the algorithm is with respect to deviations from the gamma model. To set them, the gamma model was fitted for every DSD in the database at every possible resolution from 30 s up to 30 min with a step of 30 s and the K-S test was applied to each sample. The 90th quantile of  $D_{KL}$  for all cases where the K-S test was rejected was taken as an estimate for  $C_1$ . Similarly,  $C_2$  was estimated by taking the 90th quantile of  $D_{KL}$  for all cases for which the K-S test was accepted. This resulted in  $C_1 = 0.09$  and  $C_2 = 0.05$  meaning that the tolerance level is lower when K-S test rejects than when it accepts. The values of  $C_1$  and  $C_2$  above were derived empirically without any considerations for performance. They are specific to our dataset and other values can be chosen depending on user requirements. For more discussion on the choice of  $C_1$  and  $C_2$  and how they affect the results, see appendix C.

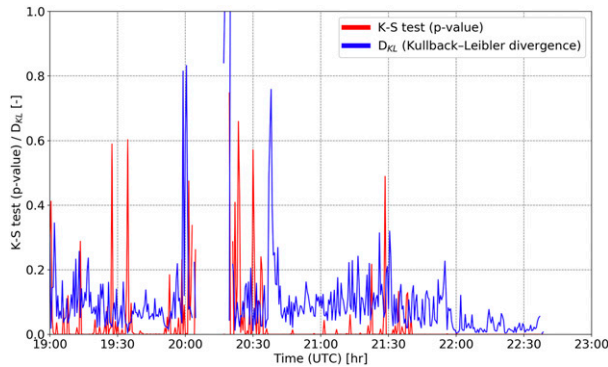


FIG. 4. The  $p$  values from K-S test and  $D_{KL}$  (Kullback–Leibler divergence) at 30-s resolution on 12 Oct 2014.

## 4. Results

### a. Adequacy of gamma model at 30 s

In the following, the adequacy of the gamma model is investigated at 30-s resolution through the K-S test and Kullback–Leibler divergence ( $D_{KL}$ ) described in sections 3c and 3d. For the following computations, the simple K-S test (without Monte Carlo simulations) and the simple MLE (without truncation and rescaling) were used. For more details about the other versions (“exact” K-S test, truncated MLE), see appendixes A and B. We start by examining the results for the study case on 12 October 2014. The  $p$  values from K-S test and  $D_{KL}$  are shown in Fig. 4. They show that 89.8% of the 30-s samples do not pass the K-S test at a significance level of  $\alpha = 0.05$ , meaning that the majority of the observed DSDs during this event are not strictly distributed according to the gamma model. At the same time, most  $D_{KL}$  values during this rain event remain relatively close to zero, indicating that the DSDs are not far away from a gamma distribution either, in contrast to what the K-S test suggests. This can be explained by the fact that  $D_{KL}$  only looks at the shape of the distribution while the power of the K-S test is heavily dependent on the number concentration, with large samples being more likely to result in rejection (Cugerone and De Michele 2015; Mohd Razali and Yap 2011).

An example of this situation can be seen from 2040 to 2100 UTC at the beginning of the second part of the event and from 2145 UTC until the end of the event (third part) where the K-S test rejects the gamma assumption but  $D_{KL}$  is low ( $D_{KL} < 0.1$ ), suggesting that the gamma model is a reasonable approximation. The disagreement between K-S test and  $D_{KL}$  during these periods was expected since the two peaks of number concentration were detected here (Fig. 1). By contrast, higher acceptance rates for the K-S test were found in the first part of the event (1900–2000 UTC) which is characterized by a relatively stable period of light rain with stationary DSD (Fig. 2) and  $p$  values that are frequently above 0.05.

Two (special) periods during which the K-S test mostly accepts the gamma hypothesis were identified from 2023 to 2037 UTC and from 2126 to 2141 UTC. The first period is characterized by a very peaked DSD covering a limited number of diameter classes (i.e., between 3 and 5) and forming a triangular

shape distribution. The second is a transitional period of very light rain between the second and the third part of the event. Both periods are characterized by small number concentrations which is the main reason the gamma model gets accepted by the K-S test, even though visually the distributions do not look like a gamma distribution, especially for the transitional period (Fig. 2).

Except for the cases when the number concentration is high enough for the  $p$  value of the K-S test to be equal to 0, we see that often a local maximum for the K-S test corresponds to a local minimum for the  $D_{KL}$ . However, the overall agreement between the two metrics (K-S test and  $D_{KL}$ ) remains weak (correlation coefficient of  $-0.31$ ), highlighting the different type of information provided by the K-S test and  $D_{KL}$ .

Based on the output of K-S test and  $D_{KL}$ , four interesting cases are presented in Fig. 5. The first (Figs. 5a,b) corresponds to a measurement made at 1934 UTC when the  $p$  value is high and the Kullback–Leibler divergence is almost zero. There is an agreement between the two metrics, since K-S test clearly accepts the gamma hypothesis and  $D_{KL}$  indicates good agreement between observations and the fitted gamma distribution (Figs. 5a,b). However, this is not always the case, as shown by the observation at 2200 UTC in Figs. 5g and 5h. Here the agreement between the observations and the fitted gamma model is visually good as confirmed by a  $D_{KL}$  value close to zero. Nevertheless, the K-S test rejects the gamma hypothesis due to a large number concentration ( $N_T = 1196 \text{ m}^{-3}$ ). The two remaining cases (Figs. 5c,d) and (Figs. 5e,f) correspond to spectra that cannot be approximated by a gamma distribution. Figures 5e and 5f show a DSD that exhibits bimodal characteristics (forcing both statistical tests to reject the gamma hypothesis no matter what the number concentration is), while Figs. 5c and 5d show a case where the Kullback–Leibler divergence is rather large ( $D_{KL} = 0.217$ ) but the K-S test accepts the gamma hypothesis ( $p$  value =  $0.287 > \alpha$ ) due to the low number concentration ( $N_T = 23 \text{ m}^{-3}$ ).

On average, over the whole event, only 1 out of 10 DSDs at 30 s strictly conformed to the gamma model (according to K-S test only). A similar low acceptance rate was found for the whole dataset. Out of 12 329 DSD spectra at 30-s resolution, 42% were accepted as gamma, of which only 21% were accepted by both K-S test and  $D_{KL}$  which means that almost 79% were not perfectly gamma according to the K-S test, but it is close enough to be approximated by one according to the  $D_{KL}$ . One reason to explain the low acceptance rate can relate to the cases with large  $N_T$  values. Another reason which could explain the high rejection of the gamma hypothesis is related to the limitations of the measurements, since Parsivel is susceptible to errors in the recorded drop concentrations, particularly for small and large drops (Raupach and Berne 2015; Tokay et al. 2014). However, the implications of this are not fully clear yet. For example, another study (Thurai and Bringi 2018) has suggested that there might be larger deviations from the gamma model at smaller diameters than suggested by the Parsivel, meaning that the rejection rates could be even higher in case better data at the low end of the spectra were available. Therefore, several other studies (Thurai and Bringi 2018; Lee et al. 2004) proposed to use the generalized gamma formulation,



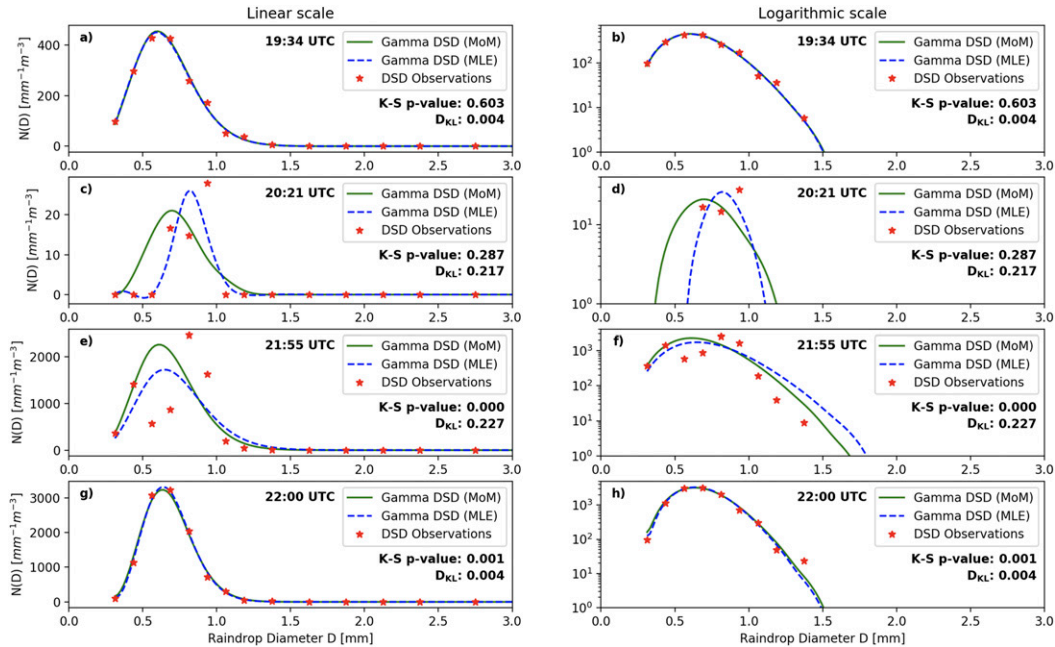


FIG. 5. Four interesting DSDs from the case study. (a),(b) High  $p$  value and low  $D_{KL}$ , (c),(d) both  $p$  value and  $D_{KL}$  relatively high, (e),(f)  $p$  value low and  $D_{KL}$  high, and (g),(h) both  $p$  value and  $D_{KL}$  low.

which is a more flexible model with additional parameters that appears to better fit naturally occurring DSDs than the standard gamma.

#### b. Influence of sampling resolution on the adequacy of the gamma model

In section 3e, a novel adaptive sampling technique was proposed where the temporal sampling resolution is adapted until the gamma model provides an adequate representation of raindrop size distributions. The approach is based on an iterative use of Kolmogorov–Smirnov goodness-of-fit test and Kullback–Leibler divergence (Fig. 3). In the following, the technique was applied to the whole dataset for 60 different sampling resolutions, starting from 30 s (sampling resolution of the original disdrometer DSD data) up to 30 min in regular steps of 30 s. If the gamma model is rejected at all resolutions, the procedure stops and the sample is flagged as being incompatible with the gamma model. Resolutions lower than 30 min were not used as the main objective is to work at the highest possible temporal resolution to capture the dynamics and microphysics of the rain and not mix different DSDs together.

In Fig. 6 the application of the decision tree algorithm to the study case is presented. It shows that for the majority of the time steps (86.8%), the gamma model was accepted at resolutions between 30 and 300 s. Most of the time (73.8%), the right part of the flowchart was followed ( $D_{KL} < C_2$ ) which means that the gamma hypothesis is not accepted by K-S test but the model is close enough to the observations to be useful. On the other hand, there are also a few interesting cases where lower resolutions are needed. One example is the period between 2048 and 2107 UTC during which the highest resolution

at which the gamma assumption is acceptable increases at every time step. Looking closer, we can see that at the beginning of that period (2045 UTC), the gamma hypothesis was accepted at rather high resolution (below 90-s resolution). However, for the rest of the period, the DSDs were consistently incompatible with the gamma model, forcing the algorithm to downsample until the beginning of the period was included.

For a deeper understanding of the adaptive sampling algorithm, some examples highlighting the sensitivity of the results to the sampling resolution are presented. The four cases which were already discussed in the previous section (section 4a, Fig. 5) are analyzed further here. Figures 7 and 8 present the fitted gamma DSDs, the  $p$  values (K-S test output) and  $D_{KL}$

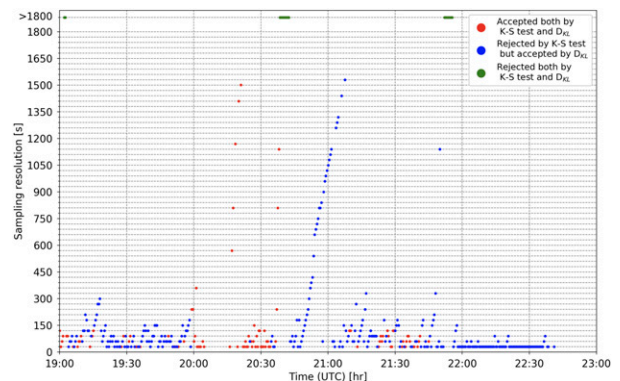


FIG. 6. The output of the adaptive sampling algorithm for determining the highest possible resolution at which the gamma hypothesis is adequate (applied on the study case on 12 Oct 2014).

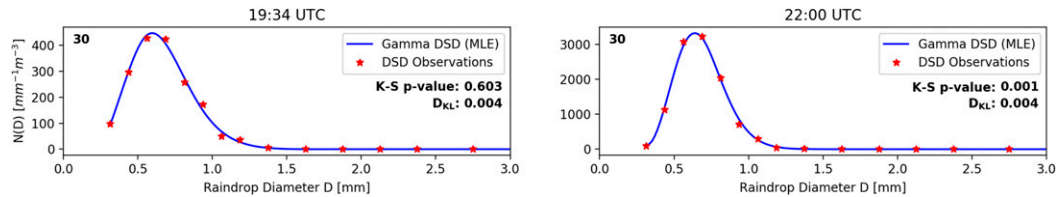


FIG. 7. DSDs at (left) 1934 and (right) 2200 UTC at 30 s. The fitted gamma distribution is shown in blue. The  $p$  values (K-S test output) and  $D_{KL}$  (Kullback–Leibler divergence) result in acceptance of the gamma hypothesis.

(Kullback–Leibler divergence) for different sampling resolutions (30, 60, 90 s, etc.) until the iterative resampling algorithm stops. Together, these four cases cover all four possible paths in the decision tree algorithm.

The first case at 1934 UTC (Fig. 7 left, Figs. 5a,b) is an example of a DSD for which the gamma model gets accepted right at the beginning at the 30-s sampling resolution (K-S test:  $0.603 > 0.05$ ,  $D_{KL}: 0.004 < 0.09$ ). Thus, there is no need to downsample. This is also the case for the spectra at 2200 UTC (Fig. 7 right, Figs. 5g,h). The only difference in this example is that the gamma hypothesis gets rejected by K-S test (K-S test:  $0.001 \not> 0.05$ ) but accepted by  $D_{KL}$  ( $D_{KL}: 0.004 < 0.05$ ) which

means that we follow the right branch of the flowchart instead of the left one. This example demonstrates the importance of the  $D_{KL}$  criterion in the algorithm when the K-S test decisions are heavily influenced by sample size. The two remaining cases are examples of situations where downsampling is required and more iterations of the algorithm are needed in order to have an acceptable agreement between the gamma model and the observations. The case at 2021 UTC (Fig. 8 left, Figs. 5c,d) is an example of a DSD that passes the K-S test ( $p$  value =  $0.287 > 0.05$ ) at 30-s resolution due to low sample size but is rejected by  $D_{KL}$  ( $0.217 \not< 0.09$ ). A lot of downsampling is required until both K-S test and  $D_{KL}$  accept the gamma assumption at 1500 s

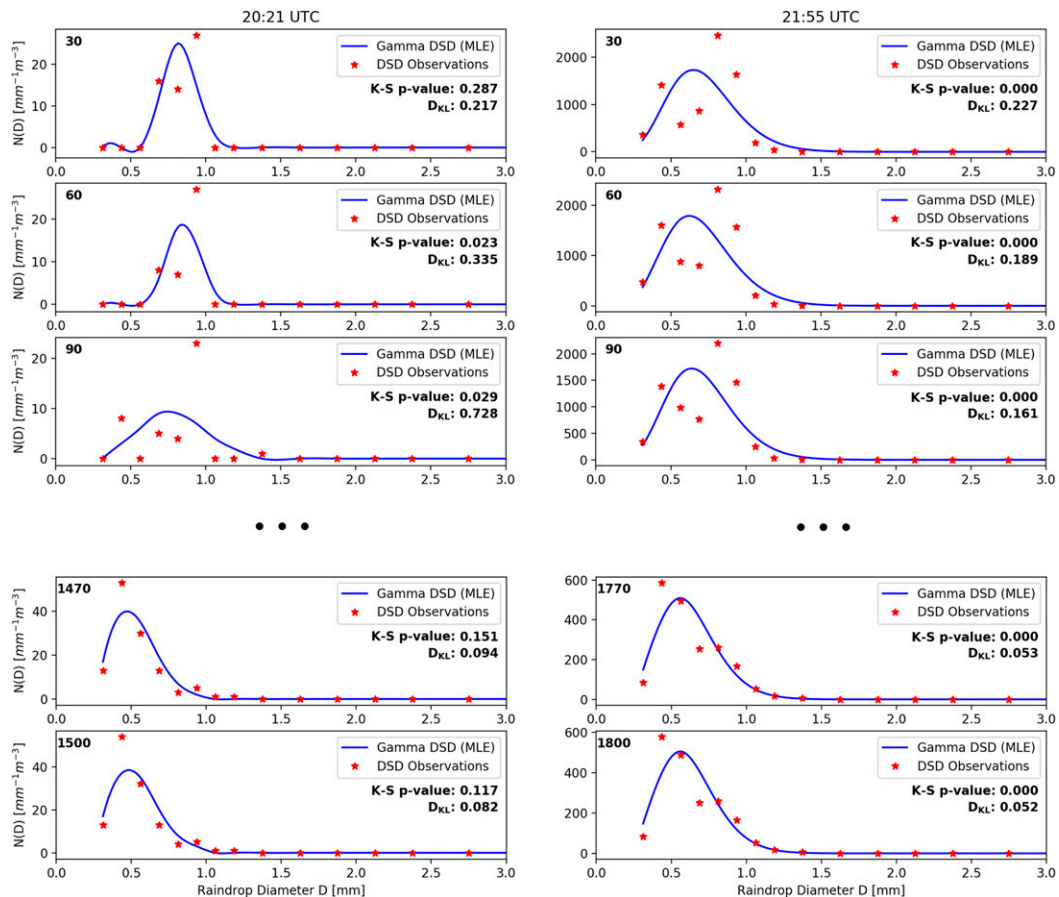


FIG. 8. DSDs at (left) 2021 and (right) 2155 UTC for different sampling resolutions (30, 60, 90 s, etc.) until the iterative resampling algorithm stops.

TABLE 2. Acceptance rates for all possible resolutions from 30 to 1800 s for the whole dataset and the relative contributions of the left and right branches of the flowchart to the acceptance rate.

Sampling resolution (s)	Acceptance rate (%)	Accepted both by K-S test and $D_{KL}$ (%)	Rejected by K-S test but accepted by $D_{KL}$ (%)
30	42.14	21.04	78.96
60	17.88	30.17	69.83
90	9.17	27.43	72.57
120	5.26	26.50	73.50
150	3.54	24.08	75.92
180	2.34	18.34	81.66
210	1.73	25.82	74.18
240	1.46	24.44	75.56
270	1.26	25.16	74.84
300	1.03	24.41	75.59
[30, 300]	85.81	24.27	75.73
[330, 600]	4.59	23.67	76.33
[630, 900]	1.75	20.37	79.63
[930, 1200]	1.22	33.11	66.89
[1230, 1500]	0.67	37.35	62.65
[1530, 1800]	0.41	33.33	66.67
>1800 (never accepted)	5.55	—	—

resolution (K-S test:  $0.117 > 0.05$ ,  $D_{KL}$ :  $0.082 < 0.09$ ). The last case (2155 UTC) is an example of a bimodal distribution. For that particular case, the distribution is too far away from the gamma model. Therefore, the algorithm rejects the gamma at all sampling resolutions up to 30 min. This last case is particularly interesting because it shows that the algorithm is also capable of revealing entire time periods during which the DSDs are completely incompatible with the gamma assumption irrespective of the temporal resolution. The latter is very valuable as mismatches between the model and the observations may not always be obvious to spot by eye just by looking at the disdrometer data. In our study case, two such periods were found (Fig. 6). They last for approximately 4 min and can be characterized as transitional rain periods corresponding to the beginning of the second and third part in the event (see section 2). Visual inspection of these two periods confirms that the DSDs are indeed bimodal. The first one (2038–2042 UTC) relates to a transition from peaked DSDs to broad DSDs with a large number of small raindrops. The second one (2152–2156 UTC) is just before the drizzle mode starts. A possible physical explanation for the second transition could be a raindrop breakup process during which larger drops break up into smaller ones. Both transitional periods are characterized by a bimodal shape resulting from the mixture of two different DSDs which is impossible to model by a gamma distribution. As long as the old and the new DSD regime overlap and none of them dominates the other, the model cannot be used.

For a more general overview of the performance of the resampling procedure, the decision tree algorithm was applied to the whole DSD dataset. Table 2 shows the acceptance rates for every possible resolution over the entire dataset. In addition, for every acceptance rate, the percentage which corresponds to the left part of the flowchart (K-S accepts and  $D_{KL} < C_1$ ) and the percentage due to the right branch of the decision tree (K-S rejects and  $D_{KL} < C_2$ ) are given. We see that the

acceptance rate of the gamma model at 30 s is 42.14%. When combined with the 60-s resolution, acceptance increases to 60.02%. When all the resolutions up to 300 s are combined, the overall acceptance rate is about 85.81%. While this is encouraging, it also means that in 15% of all cases, the gamma model did not fit the data reasonably well at high resolutions (5 min or higher). The high resolution requirement is crucial for hydrological applications or remote sensing since DSD can change quickly over the course of an event. Among the 15%, approximately one-third (5.55%) correspond to DSD spectra that do not follow a gamma distribution at any resolution (up to 1800 s), highlighting the importance of a careful DSD inspection and selection before the fitting procedure.

From the table we can also see a clear trend in the contributions of the left and right parts of the flowchart which remain stable around 25%–75% regardless of which sampling resolution was selected. This means that three out of four times the observed distribution is not perfectly gamma according to the K-S test, but it is close enough to be approximated by one according to the Kullback–Leibler divergence. Above 900 s, that proportion changes to 35%–65%; however, the cases are noticeably fewer, representing only 2.3% of the entire dataset.

At this point it is worth highlighting that these results depend on the choice of  $C_1$  and  $C_2$ , and will change for different thresholds and levels of tolerance. Even though for this study specific values for  $C_1$  and  $C_2$  were chosen, these are not optimal and other combinations can be used depending on the application. If one wants to be extremely strict,  $C_1$  and  $C_2$  can be lowered, which will result in a drop of the acceptance rates, especially at higher resolutions. In this case the algorithm becomes equivalent to a single K-S test output. On the other hand, in case  $C_1$  and  $C_2$  are increased, the algorithm will accept the vast majority of the spectra, even the ones that are very far

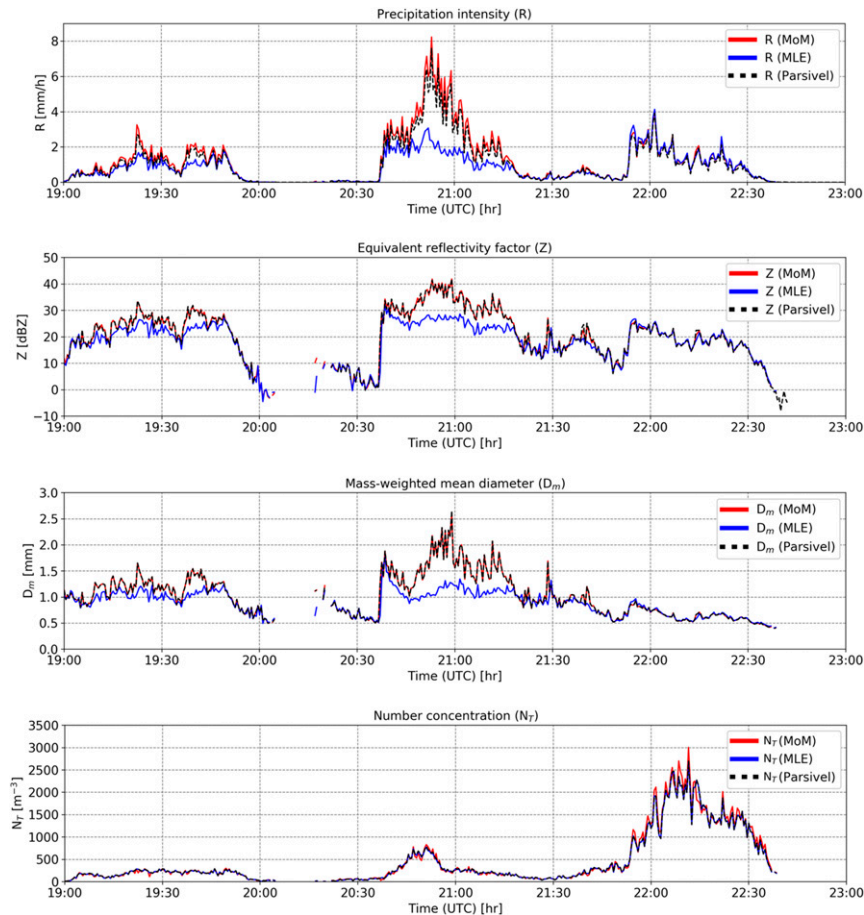


FIG. 9. Time series of (top to bottom) precipitation intensity ( $\text{mm h}^{-1}$ ), equivalent reflectivity factor (dBZ), mass-weighted mean diameter (mm), and number concentration ( $\text{m}^{-3}$ ) on 12 Oct 2014 for MoM and MLE. The black line represents the value measured by the Parsivel disdrometer.

away from a gamma distribution. In general, results are more sensitive to  $C_2$  than  $C_1$  because the former defines the level of tolerance for the right branch of the decision tree (where the most of the acceptance comes from). For more details about the influence of  $C_1$  and  $C_2$  and the sensitivity of the results to the choice of these values, see [appendix C](#).

#### c. Model adequacy based on bulk variables ( $R$ , $Z$ , $D_m$ , $N_T$ )

In the previous section, we showed how well the gamma model can describe DSD observations, based on the entire DSD spectra. However, for many applications only integrated values of the DSD ( $R$ ,  $Z$ ,  $D_m$ ,  $N_T$ ) are needed. Since these bulk variables are related to weighted moments of the DSD, the adequacy of the gamma model can also be examined as a function of these moments. For a deeper investigation of the importance of the method used to fit the DSDs, two different DSD parameter estimation methods (MoM and MLE) were used for the retrievals of  $\mu$  and  $\Lambda$ , described in [sections 3a](#) and [3b](#).

The four bulk variables corresponding to the fitted DSDs during the case study are presented in [Fig. 9](#), together with the

“true” values calculated directly from the disdrometer data. We can see that the observed bulk variables ( $N_T$ ,  $D_m$ ,  $R$ , and  $Z$ ) measured by the disdrometer and the bulk variables derived from the fitted gamma models using MLE and MoM do not always agree with each other, especially when the MLE is chosen as the parameter estimation method. This is clearly visible during the second part of the rain event (and partly during the first) during which MLE strongly underestimates  $D_m$ ,  $R$ , and  $Z$  compared with the Parsivel and MoM.

The disagreements can be explained by the fact that most DSDs during the second time period are not well approximated by the gamma model according to [Fig. 6](#). Also, the DSD spectra include bigger drops which are known to have a large influence on higher-order moments. Even though MoM makes the same mistake as MLE by assuming that the DSD is gamma, it is a better choice than MLE during this particular period because it explicitly tries to conserve the liquid water content, which is closely related to  $D_m$  and  $R$  and (to a lower degree), to the reflectivity factor  $Z$ . By contrast, MLE does not conserve the liquid water content, giving more weight to the smaller more numerous drops in the spectra. Nevertheless, one should



TABLE 3. Root-mean-square error for the four bulk variables ( $R$ ,  $Z$ ,  $D_m$ ,  $N_T$ ) at 30-s resolution for the whole dataset using method of moments (MoM) and maximum likelihood estimation (MLE).

	MoM	MLE
Rain intensity ( $R$ ) ( $\text{mm h}^{-1}$ )	0.28	0.75
Reflectivity factor ( $Z$ ) (dBZ)	0.41	3.53
Mass-weighted mean diameter ( $D_m$ ) (mm)	0.02	0.21
Number concentration ( $N_T$ ) ( $\text{m}^{-3}$ )	64.11	3.98

not discard MLE simply on that basis that it produces biased rainfall rates and reflectivities, as it can also lead to superior performance for lower-order moments of the DSD (such as  $N_T$ ). Also, it is very important to point out that MLE performs very well in cases where the DSD is in good agreement with the gamma hypothesis. This can be seen during the third part of the event during which the gamma hypothesis is reasonable and both methods (MLE and MoM) are in good overall agreement with the Parsivel observations.

Table 3 provides a more general overview of the performance of MLE and MoM for the whole dataset, showing the root-mean-square errors for four bulk variables ( $R$ ,  $Z$ ,  $D_m$ , and  $N_T$ ). We can see that MoM results in errors that are almost 3 times smaller for rain rate and 9 times for reflectivity. On the other hand, MLE performs much better than MoM for low-order moments, such as  $N_T$ . The conclusion is that both MoM and MLE can be good/bad choices depending on 1) the intended application and 2) how close or far the observed DSDs are from the gamma distribution.

Focusing on the third part of the event, the gamma DSD assumption is reasonable and the MLE method is capable of accurately estimating  $R$  and  $Z$ . This can be seen more clearly in Fig. 10 where a scatterplot between the observations of  $D_m$  and the estimations of  $D_m$  using MLE are presented. It shows that for mean diameters up to 1 mm, there is good agreement between mass-weighted mean diameter observations and model estimates regardless of the acceptance or rejection of the gamma assumption. On the other hand, for diameters above 1 mm, MLE estimates of  $D_m$  tend to be underestimated compared with the observations. This underestimation is more severe when the gamma model is rejected than when it is accepted. Consequently, the performance of MLE bulk parameters depends both on the acceptance or rejection of the gamma hypothesis and on the value of  $D_m$ .

From the last two examples (second and third parts of the rain event), it is evident that the gamma model should not be viewed as an absolute truth for all DSDs but as an approximation whose validity needs to be assessed on a case by case basis. Failing to do so can result in large errors between calculated and measured moments. The same argument applies to the DSD estimation methods based on MoM and MLE, which have different behavior and properties depending on the validity of the gamma assumption and the  $D_m$  value. For example, while MLE is superior to MoM for cases where the DSD is gamma, it should be avoided in cases where the gamma DSD assumption is questionable or  $D_m$  is large.

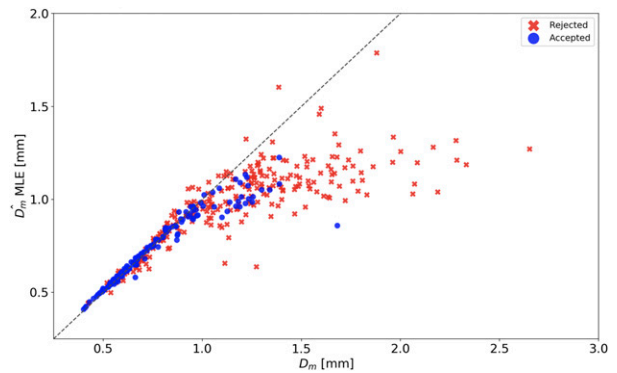


FIG. 10. Scatterplot between the observations of  $D_m$  (from the Parsivel) and the estimated values  $D_m$  using MLE when the gamma DSD hypothesis is accepted and rejected for the case study on 12 Oct 2014.

Similarly, in case the DSD is not perfectly gamma, it is better not to use truncated and rescaled MLE (for more details see appendix A). The adaptive sampling algorithm presented in this paper can provide information about the adequacy of the gamma model from a statistical point of view, which can be helpful for interpreting changes in the microphysics of rain and selecting the most appropriate fitting method. This is an often overlooked aspect of DSD analyses that is crucial for creating robust and representative DSD databases for use in radar retrievals.

## 5. Conclusions

A critical evaluation of the adequacy of the gamma model for representing raindrop size distributions was presented. The results are based on DSD data collected by a Parsivel optical disdrometer during a 2-month campaign in the Netherlands. A study case was presented and a table summarizing the results for the whole DSD dataset was provided. At first, the adequacy of the gamma model at 30 s was analyzed using the K-S test and Kullback–Leibler divergence and four interesting cases were highlighted. Then, the influence of the sampling resolution on the adequacy of the gamma model was investigated. A novel adaptive sampling technique was proposed to determine the highest temporal sampling resolution at which the gamma model provides an adequate representation of sampled DSDs. Finally, in order to assess the gamma DSD model from a more practical point of view, the accuracy of retrieved bulk variables ( $R$ ,  $Z$ ,  $D_m$ ,  $N_T$ ) was examined. According to the results the following conclusions can be drawn.

- 1) The majority of the DSD spectra are not perfectly gamma but are well approximated by the gamma model at high sampling resolutions (60.02% of the DSDs up to 60 s, 85.81% of the DSDs up to 300 s). However, a substantial number of DSDs (5.55%) were absolutely not complying with the gamma model, which means that careful selection of the DSD spectra is needed before fitting.
- 2) About three out of four times (across all temporal scales), the gamma hypothesis gets accepted not because of the



TABLE A1. Root-mean-square error for the four bulk variables ( $R$ ,  $Z$ ,  $D_m$ ,  $N_T$ ) at 30-s resolution, for the whole DSD dataset, for the nongamma DSDs, and for gamma DSDs using MLE for truncated and binned data.

	All DSDs	Nongamma DSDs	Gamma DSDs
Rain intensity ( $R$ ) ( $\text{mm h}^{-1}$ )	0.89	0.97	0.37
Reflectivity factor ( $Z$ ) (dBZ)	7.41	7.81	2.63
Mass-weighted mean diameter ( $D_m$ ) (mm)	0.31	0.34	0.12
Number concentration ( $N_T$ ) ( $\text{m}^{-3}$ )	22.27	50.79	9.14

K-S test (it is not perfectly gamma) but because of the Kullback–Leibler divergence. Thus, most DSDs are not truly gamma but come close to it.

- 3) One should not fit the two parameters  $\mu$  and  $\Lambda$  of the DSD using MLE when the distribution is not perfectly gamma or, at least, very close to it. The same argumentation holds for truncated and rescaled MLE. Failing to do so results in underestimated  $D_m$  and  $R$  values. In this case the safer option is to use the MoM assuming that we want to retrieve high-order moments of the distribution (e.g., rain intensity, reflectivity factor).
- 4) The adaptive sampling algorithm proposed in this paper is capable of automatically identifying transition periods during which the DSD cannot be represented by a gamma model (at any resolution). These may not be easily visible in the data but become very clear after applying our algorithm.

Finally, it should be mentioned that this study is not a statement against the use of the gamma model which is often a good approximation. However, it highlights the importance of checking the adequacy of these assumptions. It also lays the foundation for a better automatic quality control of DSD retrievals for use in remote sensing applications. The main idea could also be applicable to other relevant research in the future, including an evaluation of the gamma model assumption on  $\mu$ – $\Lambda$  relationships used in polarimetric radar retrievals.

**Acknowledgments.** This work was supported by the Netherlands Organisation for Scientific Research (NWO) through the “User Support Programme Space Research 2012–2016,” Project ALW-GO/15-35. We are grateful to TROPOS for collecting and sharing Parsivel disdrometer data used in this study.

## APPENDIX A

### MLE for Truncated and Binned Data

The rescaling of the likelihood function in Eq. (13) and the theoretical complications that follow from it are heavily dependent on the assumption that the DSD is indeed a gamma distribution, which we know is hardly ever the case. As a result, the corrections are not necessarily beneficial and could

actually make the fit worse. To quantify this, we applied the rescaling and truncation to the case study. With this new way of fitting, we saw a drastic decrease in the cumulative acceptance rate up to 5 min from 86.8% to 65.3%. This can be explained if we take into account that most of the spectra of the case study at 30-s resolution are not gamma (according to K-S test only 10% are accepted). Consequently, the correction (rescaling) for the remaining 90% of them was based on the wrong initial hypothesis which led to a worse MLE fit than before.

To investigate this issue further, we applied the new rescaled and truncated MLE fit to the entire DSD dataset at 30 s resolution and we calculated the 4 bulk variables ( $R$ ,  $Z$ ,  $D_m$ ,  $N_T$ ) corresponding to the fitted DSDs. Combining them together with the “true” values calculated directly from the disdrometer, we derived the root-mean-square error (RMSE) of the bulk variables for the whole campaign (Table A1). The results show that the RMSE of each variable calculated using the truncated and rescaled MLE increased compared with the values presented in Table 3 (MLE without rescaling). The RMSE of the rain intensity increased from 0.75 to 0.89  $\text{mm h}^{-1}$ , reflectivity from 3.53 to 7.41 dBZ,  $D_m$  from 0.21 to 0.31 mm, and  $N_T$  from 3.98 to 22.27  $\text{m}^{-3}$ . This is attributed to the fact that most of the time, the DSDs are not strictly gamma. This can be seen very clearly if we calculate the RMSE of the bulk variables separately for all the DSDs which satisfy the gamma assumption (according to our algorithm) and for those which do not (Table A1). For the gamma shaped DSDs, the truncation and rescaling resulted in lower RMSE values from 0.89 to 0.37  $\text{mm h}^{-1}$  for rain intensity, from 7.41 to 2.63 dBZ for reflectivity, from 0.31 to 0.12 mm for mean diameter, and from 22.27 to 9.14  $\text{m}^{-3}$  for number concentration. On the other hand, for the nongamma DSDs the RMSE increased to 0.97  $\text{mm h}^{-1}$  for rain intensity, 7.81 dBZ for reflectivity, 0.34 mm for mean diameter, and 50.79  $\text{m}^{-3}$  for the number concentration.

All the above shows that the best way to estimate the parameters when we are not sure whether the distribution is really gamma or not is to use

- estimates which are based on a few moments calculated directly from the sample (see Table 3, RMSE MoM), instead of fitting the whole density function as in MLE;
- simple numerical solutions which make as little assumptions as possible about the underlying distribution (MLE without rescaling and truncation).

## APPENDIX B

### “Exact” K-S Test (Using Monte Carlo Simulations)

An alternative way to apply the K-S test, suitable for cases when population parameters are unknown and must be estimated by sample statistics was implemented. Specifically, we applied the LcKS function from the KScorrect package (R programming language), which uses Monte Carlo simulations to estimate the  $p$  values. A total of 4999 random samples (recommended by the authors of the package) were drawn from a gamma distribution with parameters calculated from

TABLE C1. Sensitivity analysis of the thresholds  $C_1$  and  $C_2$  (case study only). The impact of different  $C_1$  and  $C_2$  on the cumulative acceptance rates up to 5 min of the gamma DSD hypothesis is assessed. The relative contributions of the left and right branches of the flowchart are also given. The first combination of  $C_1$  and  $C_2$  values ( $C_1 = 0.09$  and  $C_2 = 0.05$ ) is the one used in this study.

$C_1/C_2$	Acceptance rate (%)	Accepted both by K-S test and $D_{KL}$ (%)	Rejected by K-S test but accepted by $D_{KL}$ (%)
0.09/0.05	86.83	25.12	74.88
0.18/0.05	87.07	27.73	72.27
0.045/0.05	85.12	23.06	76.94
0.09/0.1	94.88	14.91	85.09
0.09/0.025	66.59	57.51	42.49

the sample. Based on these simulations we found that there were no significant changes to the final results. For the study case, the acceptance rates at resolutions between 30 and 300 s differed by less than 1% compared with the previous ones, increasing from 86.8% to 87.3%. Also, because the “exact” K-S test resulted in a lower rejection rate, the left branch of the decision tree was followed slightly more often (from 25%/75% it went to 28%/72%). But overall, no big differences could be observed. This is due to the algorithm construction itself, which is mostly based on the value of  $D_{KL}$ . The purpose of the K-S test is mainly to give a first opinion, but the final decision is always based on  $D_{KL}$ .

Because there is no significant difference in term of acceptance/rejection rates and because the “exact” K-S test is computationally very expensive and slow, we recommend using the simpler, slightly biased version of the K-S test without Monte Carlo simulations when applying the resampling algorithm.

## APPENDIX C

### Sensitivity Analysis for $C_1$ and $C_2$

A sensitivity analysis was carried out in order to investigate the stability of the acceptance and rejection rates of the gamma hypothesis for different  $C_1$  and  $C_2$  values. A total of four new combinations of  $C_1$  and  $C_2$  values were considered, corresponding to twice/half the original values (see Table C1). For each combination, the adaptive resampling algorithm was applied at every possible resolution from 30 s up to 5 min with a step of 30 s. The cumulative acceptance rates (up to 5 min) and the contributions of the two branches were derived (Table C1). Table C1 shows that when  $C_2$  is fixed ( $C_2 = 0.05$ ) and we change  $C_1$  (from 0.09 to either 0.18 or 0.045), the results remain relatively stable. The acceptance rate varies by  $\pm 1\%$  and the left and right branch contributions of the decision tree by  $\pm 2\%$ . When  $C_1$  is fixed ( $C_1 = 0.09$ ) and  $C_2$  takes different values, the results are more sensitive. When we increase the  $D_{KL}$  tolerance ( $C_2$ ) from 0.05 to 0.1, almost 95% of the cases get accepted, mainly through the right branch of the decision tree ( $\sim 15/85$ ). However, it should be noted that this specific combination of values ( $C_1 = 0.09$  and  $C_2 = 0.1$ ) is not realistic since by definition  $C_1$  should be greater than  $C_2$  to ensure a

lower tolerance on  $D_{KL}$  in case the K-S test gets rejected. As for the last case where the tolerance on  $D_{KL}$  is much lower ( $C_2 = 0.025$ ), we see a drastic drop in the acceptance rate to 66%.

Overall, from the sensitivity analysis we conclude that

- values for  $C_1$  and  $C_2$  can be adapted depending on user requirements and application,
- the algorithm is more sensitive on  $C_2$ , and
- results depend on the choice of  $C_1$  and  $C_2$ . The chosen values for this study are not claimed to be optimal.

## REFERENCES

- Adirosi, E., E. Gorgucci, L. Baldini, and A. Tokay, 2014: Evaluation of gamma raindrop size distribution assumption through comparison of rain rates of measured and radar-equivalent gamma DSD. *J. Appl. Meteor. Climatol.*, **53**, 1618–1635, <https://doi.org/10.1175/JAMC-D-13-0150.1>.
- , L. Baldini, F. Lombardo, F. Russo, F. Napolitano, E. Volpi, and A. Tokay, 2015: Comparison of different fittings of drop spectra for rainfall retrievals. *Adv. Water Resour.*, **83**, 55–67, <https://doi.org/10.1016/j.advwatres.2015.05.009>.
- , E. Volpi, F. Lombardo, and L. Baldini, 2016: Raindrop size distribution: Fitting performance of common theoretical models. *Adv. Water Resour.*, **96**, 290–305, <https://doi.org/10.1016/j.advwatres.2016.07.010>.
- Bringi, V. N., V. Chandrasekar, J. Hubbert, E. Gorgucci, W. L. Randeu, and M. Schoenhuber, 2003: Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis. *J. Atmos. Sci.*, **60**, 354–365, [https://doi.org/10.1175/1520-0469\(2003\)060<0354:RSDIDC>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<0354:RSDIDC>2.0.CO;2).
- Chambers, J., W. Cleveland, B. Kleiner, and P. Tukey, 1983: *Graphical Methods For Data Analysis*. Duxbury, 410 pp.
- Cugeron, K., and C. De Michele, 2015: Johnson sb as general functional form for raindrop size distribution. *Water Resour. Res.*, **51**, 6276–6289, <https://doi.org/10.1002/2014WR016484>.
- Ekerete, K.-M. E., F. H. Hunt, J. L. Jeffery, and I. E. Otung, 2015: Modeling rainfall drop size distribution in southern England using a Gaussian mixture model. *Radio Sci.*, **50**, 876–885, <https://doi.org/10.1002/2015RS005674>.
- Gires, A., I. Tchiguirinskaia, D. Schertzer, and A. Berne, 2015: 2DVD data revisited: Multifractal insights into cuts of the spatiotemporal rainfall process. *J. Hydrometeorol.*, **16**, 548–562, <https://doi.org/10.1175/JHM-D-14-0127.1>.
- Ignaccolo, M., and C. De Michele, 2014: Phase space parameterization of rain: The inadequacy of gamma distribution. *J. Appl. Meteor. Climatol.*, **53**, 548–562, <https://doi.org/10.1175/JAMC-D-13-050.1>.
- , —, and S. Bianco, 2009: The droplike nature of rain and its invariant statistical properties. *J. Hydrometeorol.*, **10**, 79–95, <https://doi.org/10.1175/2008JHM975.1>.
- Jaffrain, J., and A. Berne, 2011: Experimental quantification of the sampling uncertainty associated with measurements from Parsivel disdrometers. *J. Hydrometeorol.*, **12**, 352–370, <https://doi.org/10.1175/2010JHM1244.1>.
- Jameson, A. R., and A. B. Kostinski, 2001: What is a raindrop size distribution? *Bull. Amer. Meteor. Soc.*, **82**, 1169–1178, [https://doi.org/10.1175/1520-0477\(2001\)082<1169:WIARSD>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1169:WIARSD>2.3.CO;2).
- Johnson, R. W., D. V. Kliche, and P. L. Smith, 2011: Comparison of estimators for parameters of gamma distributions with left-truncated samples. *J. Appl. Meteor. Climatol.*, **50**, 296–310, <https://doi.org/10.1175/2010JAMC2478.1>.

- , —, and —, 2014: Maximum likelihood estimation of gamma parameters for coarsely binned and truncated raindrop size data. *Quart. J. Roy. Meteor. Soc.*, **140**, 1245–1256, <https://doi.org/10.1002/qj.2209>.
- , —, and —, 2015: Modeling raindrop size. *J. Stat. Educ.*, **23**, 1–26, <https://doi.org/10.1080/10691898.2015.11889725>.
- Kliche, D. V., P. L. Smith, and R. W. Johnson, 2008: L-moment estimators as applied to gamma drop size distributions. *J. Appl. Meteor. Climatol.*, **47**, 3117–3130, <https://doi.org/10.1175/2008JAMC1936.1>.
- Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86, <https://doi.org/10.1214/aoms/1177729694>.
- Laio, F., 2004: Cramer–von Mises and Anderson–Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour. Res.*, **40**, W09308, <https://doi.org/10.1029/2004WR003204>.
- Lee, G. W., I. Zawadzki, W. Szyrmer, D. Sempere-Torres, and R. Uijlenhoet, 2004: A general approach to double-moment normalization of drop size distributions. *J. Appl. Meteor.*, **43**, 264–281, [https://doi.org/10.1175/1520-0450\(2004\)043<0264:AGATDN>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0264:AGATDN>2.0.CO;2).
- Löffler-Mang, M., and J. Joss, 2000: An optical disdrometer for measuring size and velocity of hydrometeors. *J. Atmos. Oceanic Technol.*, **17**, 130–139, [https://doi.org/10.1175/1520-0426\(2000\)017<0130:AODFMS>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<0130:AODFMS>2.0.CO;2).
- Marshall, J. S., and W. M. K. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165–166, [https://doi.org/10.1175/1520-0469\(1948\)005<0165:TDORWS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2).
- Mohd Razali, N., and B. Yap, 2011: Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *J. Stat. Model. Analytics*, **2**, 21–33.
- Ralph, F. M., and Coauthors, 2014: A vision for future observations for western U.S. extreme precipitation and flooding. *J. Contemp. Water Res. Educ.*, **153**, 16–32, <https://doi.org/10.1111/j.1936-704X.2014.03176.x>.
- Raupach, T. H., and A. Berne, 2015: Correction of raindrop size distributions measured by Parsivel disdrometers, using a two-dimensional video disdrometer as a reference. *Atmos. Meas. Tech.*, **8**, 343–365, <https://doi.org/10.5194/amt-8-343-2015>.
- Rose, C. R., and V. Chandrasekar, 2006: A GPM dual-frequency retrieval algorithm: DSD profile-optimization method. *J. Atmos. Oceanic Technol.*, **23**, 1372–1383, <https://doi.org/10.1175/JTECH1921.1>.
- Schleiss, M. A., A. Berne, and R. Uijlenhoet, 2009: Geostatistical simulation of two-dimensional fields of raindrop size distributions at the meso- $\gamma$  scale. *Water Resour. Res.*, **45**, W07415, <https://doi.org/10.1029/2008WR007545>.
- , S. Chamoun, and A. Berne, 2014: Nonstationarity in intermittent rainfall: The “dry drift.” *J. Hydrometeorol.*, **15**, 1189–1204, <https://doi.org/10.1175/JHM-D-13-095.1>.
- Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of “normalized” distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing. *J. Appl. Meteor.*, **40**, 1118–1140, [https://doi.org/10.1175/1520-0450\(2001\)040<1118:TCOND>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1118:TCOND>2.0.CO;2).
- Thurai, M., and V. N. Bringi, 2018: Application of the generalized gamma model to represent the full rain drop size distribution spectra. *J. Appl. Meteor. Climatol.*, **57**, 1197–1210, <https://doi.org/10.1175/JAMC-D-17-0235.1>.
- , C. Williams, and V. Bringi, 2014: Examining the correlations between drop size distribution parameters using data from two side-by-side 2D-video disdrometers. *Atmos. Res.*, **144**, 95–110, <https://doi.org/10.1016/j.atmosres.2014.01.002>.
- Tokay, A., D. B. Wolff, and W. A. Petersen, 2014: Evaluation of the new version of the laser-optical disdrometer, OTT Parsivel<sup>2</sup>. *J. Atmos. Oceanic Technol.*, **31**, 1276–1288, <https://doi.org/10.1175/JTECH-D-13-00174.1>.
- Uijlenhoet, R., M. Steiner, and J. A. Smith, 2003: Variability of raindrop size distributions in a squall line and implications for radar rainfall estimation. *J. Hydrometeorol.*, **4**, 43–61, [https://doi.org/10.1175/1525-7541\(2003\)004<0043:VORSDI>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0043:VORSDI>2.0.CO;2).
- Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. *J. Climate Appl. Meteor.*, **22**, 1764–1775, [https://doi.org/10.1175/1520-0450\(1983\)022<1764:NVITAF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1764:NVITAF>2.0.CO;2).
- , 1985: The effects of drop size distribution truncation on rainfall integral parameters and empirical relations. *J. Climate Appl. Meteor.*, **24**, 580–590, [https://doi.org/10.1175/1520-0450\(1985\)024<0580:TEODSD>2.0.CO;2](https://doi.org/10.1175/1520-0450(1985)024<0580:TEODSD>2.0.CO;2).
- Watanabe, Y., and D. M. Ingram, 2016: Size distributions of sprays produced by violent wave impacts on vertical sea walls. *Proc. Roy. Soc. London*, **472A**, 20160423, <https://doi.org/10.1098/rspa.2016.0423>.
- Yakubu, M. L., Z. Yusop, and F. Yusof, 2014: The modelled raindrop size distribution of Skudai, peninsular Malaysia, using exponential and lognormal distributions. *Sci. World J.*, **2014**, 361703, <https://doi.org/10.1155/2014/361703>.
- Zhang, G., J. Vivekanandan, and E. Brandes, 2001: A method for estimating rain rate and drop size distribution from polarimetric radar measurements. *IEEE Trans. Geosci. Remote Sens.*, **39**, 830–841, <https://doi.org/10.1109/36.917906>.