Comparison of different fittings of experimental DSD

Elisa Adirosi (1,2), Luca Baldini (1), Federico Lombardo (3), Ali Tokay (4,5), Fabio Russo (2), and Francesco Napolitano (2)

(1) Institute of Atmospheric Sciences and Climate (CNR), Rome, Italy (2) Department of Civil, Building and Environmental Engineering, Sapienza University of Rome, Rome, Italy
(3) Department of Engineering, Roma Tre University, Rome, Italy (4) Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, Maryland , (5) NASA Goddard Space Flight Center, Greenbelt, Maryland

Motivations

- Raindrop size distributions (DSD) can be written as $N(D)=n_c f(D)$: n_c is the number concentration, $f(D)$ a pdf
- Several pdfs have been proposed: the gamma is the most frequently assumed [Adirosi et al. 2014].
- The **third to sixth moments** of the DSD are proportional to relevant **hydrological and meteorological parameters** (i.e. rain rate, liquid water content, kinetic energy, reflectivity): performance of retrievals depends on the form of DSD assumed.
- Higher the order of moments, the **higher is the weight of the drops in the DSD right tail**.
- The form of the right tail of a DSD governs both the magnitude and the frequency of the bigger drops.
- Retrieval of parametric DSDs aims to model the whole drop spectra: the retrieved DSD could not be adequate for specific portions, such as the upper tail, critical for some meteorological parameters.
- The **method of moments,** commonly used for retrieving DSD parameters, is **known to be biased** [Smith and Kliche 2005] determining unrealistic variability of retrieved parameters [Ulbrich and Atlas, 1998].

Objectives

To investigate the performance of four one-sided continuous distributions with different tails (the **Pareto**, the **Lognormal**, the **Gamma** and the **Weibull** distributions) through fitting both

- to the large drops only (**upper tail fitting**)
- to the entire measured spectra (**whole fitting**)

Heavier tail → Lighter tail

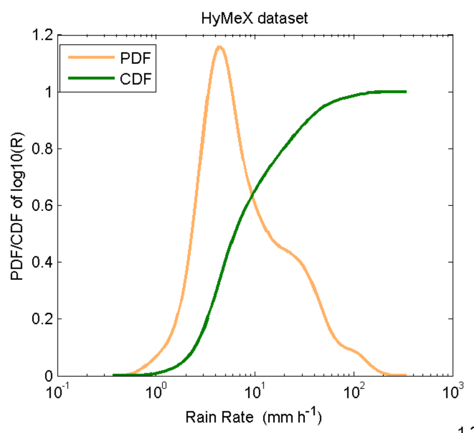
Pareto	Lognormal	Weibull γ<1	Gamma	Weibull γ>1

Measured Datasets

Data consist of 1-min drop spectra collected during **HyMeX SOP1.1 (Rome)** and **MC3E (Central Oklahoma)** by a NASA two-dimensional video disdrometer (**2DVD**) provided to support GPM pre-launch ground validation activities.

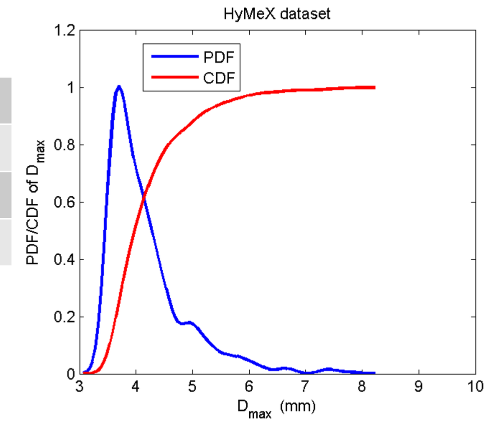
Only the **spectra with a minimum of 100 drops and D_{max} > 3.5 mm** are retained: large drops are adequately sampled, but DSDs of light rain can be discarded.

HyMex (485 1-min DSDs)

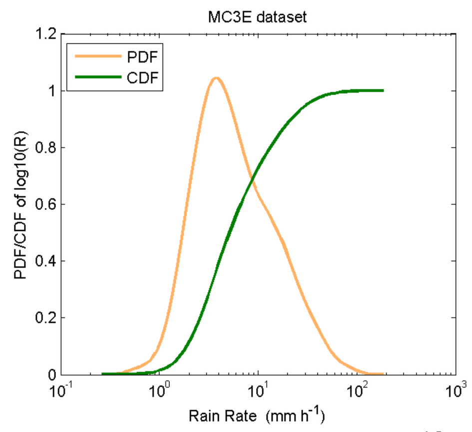


	R (mm h⁻¹)
min	0.83
mean	13.65
max	150.13

	D_{max} (mm)
min	3.51
mean	4.18
max	7.79

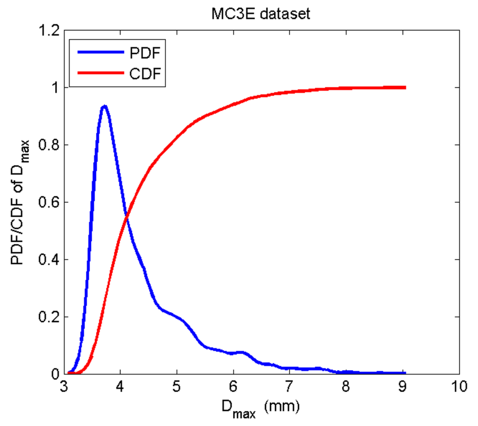


MC3E (975 1-min DSDs)



	R (mm h⁻¹)
min	0.53
mean	9.06
max	90.15

	D_{max} (mm)
min	3.51
mean	4.31
max	8.61



Tail definition: upper part of the drop spectrum with D > D_{th} → 90th percentile of drop diameters measured in a given minute

Fitting Method

The fitting method permits to directly fit any theoretical distribution to the empirical one minimizing numerically the modified mean square error (MSE) norm *N1* [Papalexiou at al., 2013]:

$$N1=\frac{1}{N}\sum_{i=n-N+1}^{n}\left(\frac{\overline{F_T}(D_i)}{\overline{F_E}(D_i)}-1\right)^2$$

n = is the total number of drops measured in 1 minute
- Tail Fitting *N* = number of drops with $D \geq D_{th}$
- Whole Fitting *N* = *n*

$\overline{F_E}(D_i)$ is the **empirical** exceedance probability function (EPF). First the 1-min *N(D)* is computed from the 2DVD raw data

$$N(D_i)=\frac{1}{\Delta t}\sum_{k=1}^{M_i}\frac{1}{A_k v_k}$$

where Δt (s) is the time interval, M_i is the total number of drops in a given minute, v_k (m s^{-1}) is the fall velocity and A_k (mm^2) is the cross-sectional area.
Then the probability density function of drop diameter (f_D) is computed and the empirical EPF can be obtained. Note that when the three parameter gamma distribution is used to model the measured DSD

$f_D(D)=\frac{\Lambda^{\mu+1}}{\Gamma(\mu+1)}e^{-\Lambda D}D^{\mu}$; μ = shape parameter and Λ (mm^{-1}) = slope parameter

$\overline{F_T}(D_i)$ is the **theoretical** exceedance probability function (EPF) computed considering each single drop measured in a given minute:

Pareto **(PA)** → $\overline{F}_{PA}(x)=\left(1+\gamma\left(\frac{x}{\beta}\right)\right)^{-\frac{1}{\gamma}}$ $with\ \beta>0\ and\ \gamma\geq 0$

Lognormal **(LN)** → $\overline{F}_{LN}(x)=\frac{1}{2}\,erfc\left(ln\left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma}}\right)$; $with\ \beta>0\ and\ \gamma>0$

Weibull **(WE)** → $\overline{F}_{WE}(x)=\exp\left(-\left(\frac{x}{\beta}\right)^{\gamma}\right)$ $with\ \beta>0\ and\ \gamma>0$

Gamma **(GA)** → $\overline{F}_{GA}(x)=\Gamma\left(\gamma,\frac{x}{\beta}\right)/\Gamma(\gamma)$;

$$\Gamma(s,x)=\int_x^{\infty}t^{s-1}e^{-t}\,dt;\quad \Gamma(s)=\int_0^{\infty}t^{s-1}e^{-t}\,dt;$$

where β and γ are the scale and shape parameter, respectively. The shape parameter controls the asymptotic behavior of the tail.

Results

Whole Fitting

Although the two datasets were collected in two different areas of the world, the estimated parameters β and γ are similar. The shape parameter (γ) of Pareto distribution is very low (means equal to 0.01): the Pareto degenerates to the exponential tail (γ = 0). Furthermore its performance is the worst (i.e. *N1* values are the highest).

LN, WE, and GA fittings were compared in pair:

HYMEX DATASET			
	LN	WE	GA
LN	\	38.4%	30.5%
WE	61.6%	\	43.9%
GA	69.5%	56.1%	\

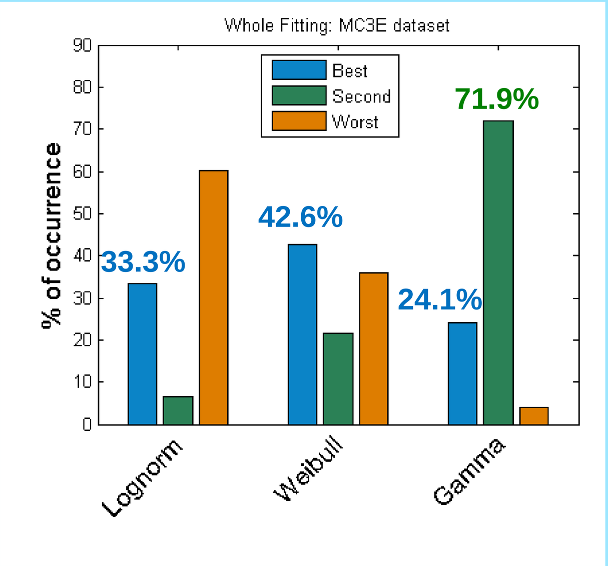
MC3E DATASET			
	LN	WE	GA
LN	\	39.5%	33.6%
WE	60.5%	\	46.3%
GA	66.4%	53.7%	\

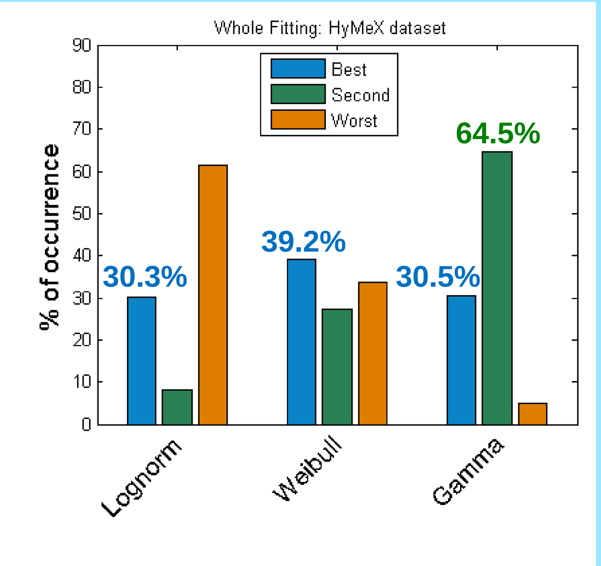
The gamma distribution was better fitted in about 60% of the 1-min measured spectra.

Note that the μ and Λ parameters con be obtained from the estimated β and γ as:
- $\mu = \gamma - 1$
- $\Lambda = 1/\beta$.

PERFORMANCE OF THE FITTINGS

The distribution that better fits a given drop spectrum is the one with the minimum *N1*.





Despite the Weibull distribution is the one with the highest percentages of success, it is closely followed by the Gamma distribution. In fact when the Gamma is not the best, its *N1* values are slightly bigger than the ones of the Weibull.

Upper tail Fitting

As for the whole fitting, the Pareto distribution degenerates to exponential (shape parameter ~ 0); the mean and the median of *N1* are the highest, with a percentage of success around 1%. These findings confirm that a **heavy tail is not adequate** to model the upper part of the disdrometer measured drop spectra.

LN, WE and GA fittings compared in pair:

HYMEX DATASET			
	LN	WE	GA
LN	\	40.6%	39.2%
WE	59.4%	\	49.7%
GA	60.8%	50.3%	\

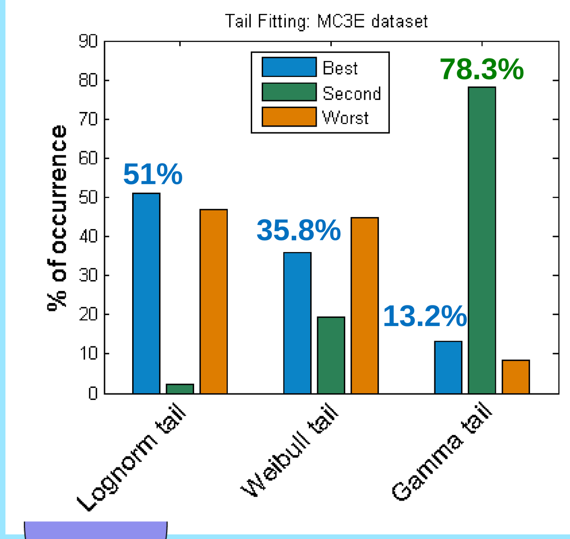
MC3E DATASET			
	LN	WE	GA
LN	\	52.9%	51.3%
WE	47.1%	\	44.0%
GA	48.7%	56.8%	\

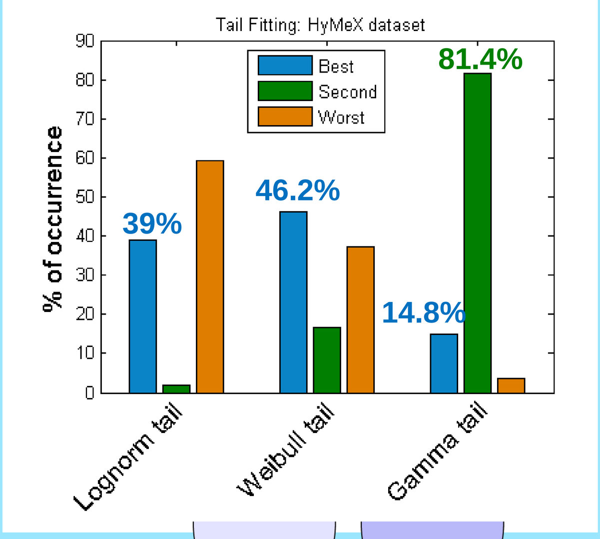
PERFORMANCE OF THE FITTINGS

When the tail fitting is performed:
- the Lognormal distribution is the best fit for MC3E
- the Weibull distribution is the best fit for HyMex
- the percentages of spectra better fitted by the Gamma distribution decrease with respect to the whole fitting
- when the Gamma tail distribution is not the best, it is the second one with a very high percentage of success

The results suggest that the **most of the disdrometer measured drop spectra** can be described with a **light-tailed distribution.**

The **practical implication** is that in a light-tailed distribution the larger drops are predicted less frequent than in a heavy-tailed distribution.





Heavy tail | Very light tail | Light tail

Note that when the Weibull distribution performs better the shape parameters are greater than 1.

References: - Smith, P. L., and D. Kliche, *J. Appl. Meteor.*, **44**, 1195-1205 (2005) - Ulbrich, C. W., and Atlas D., *J. Appl. Meteor.*, **37**, 912-923 (1998) - Papalexiou, S. P., Koutsoyiannis, D., Makropoulos, C., *Hydrol. Earth Syst. Sci.*, **17**, 851-862 (2013) - Adirosi, E., L. Baldini, F. Lombardo, F. Russo and F. Napolitano, *AIP Conf. Proc.*, **1558**, 1669-1672 (2013) - Adirosi E., Gorgucci E., Baldini L. , Tokay A., *J. Appl. Meteorol. Clim.*, in press, (2014) doi: http://dx.doi.org/10.1175/JAMC-D-13-0150.1