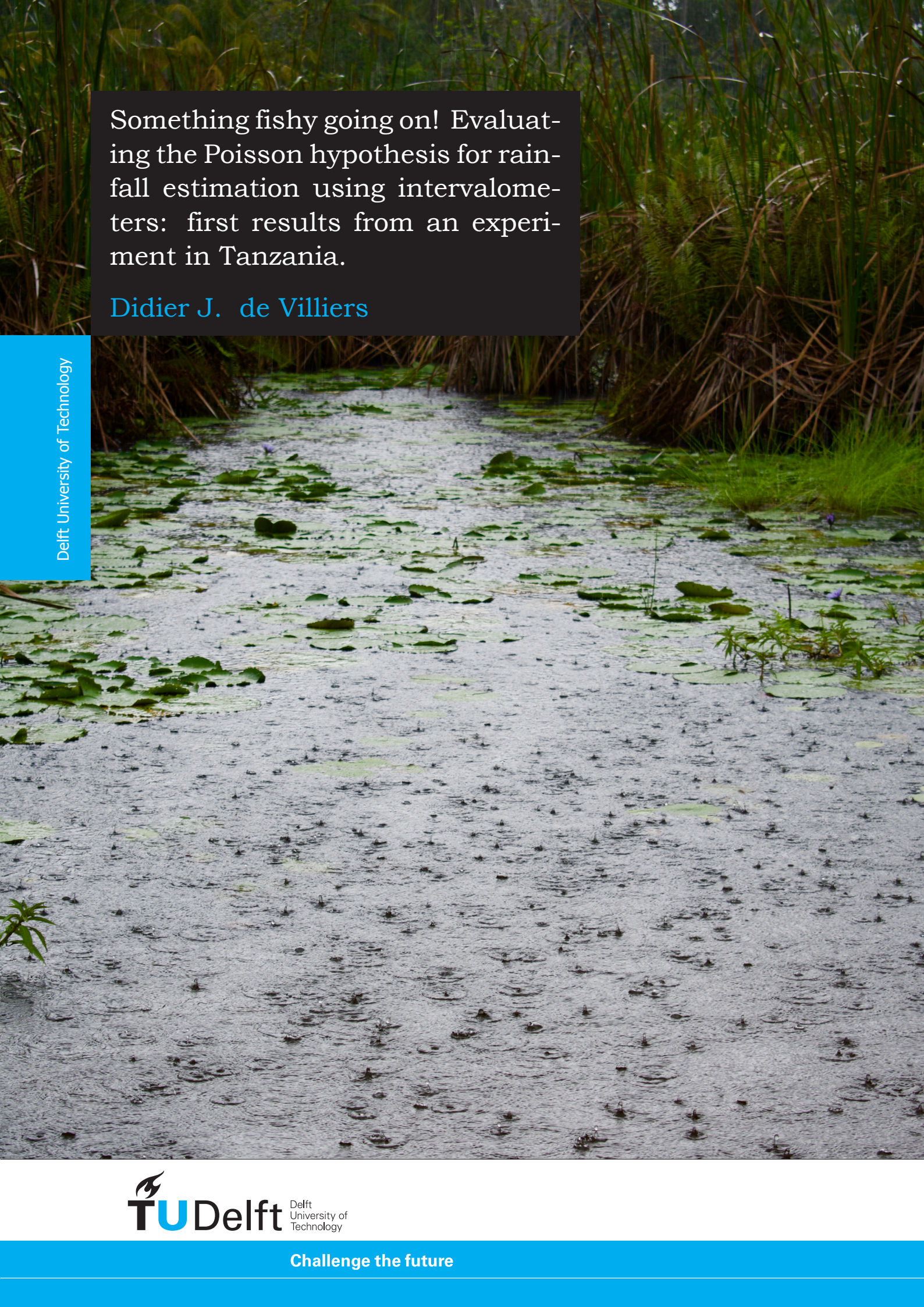




Something fishy going on! Evaluating the Poisson hypothesis for rainfall estimation using intervalometers: first results from an experiment in Tanzania.

Didier J. de Villiers

Delft University of Technology

Something fishy going on! Evaluating the Poisson hypothesis for rainfall estimation using intervalometers: first results from an experiment in Tanzania.

by

Didier J. de Villiers

in partial fulfillment of the requirements for the degree of

Master of Science

in Environmental Engineering

at the Delft University of Technology,

to be defended publicly on Wednesday August 21, 2019 at 14:00.

Supervisor:	Prof. dr. ir. Nick van de Giesen	
Thesis committee:	Prof. dr. ir. Nick van de Giesen,	TU Delft
	Dr. Ir. Marie-Claire ten Veldhuis,	TU Delft
	Dr. Marc Schleiss,	TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Acknowledgements

No project is a solitary endeavour and this one is certainly no exception. Many people have helped me over the course of this research in many different ways by providing opinions and comments, time and energy, making resources available or giving me support when I have struggled. I would like to thank you all and I apologise if you are not mentioned directly by name.

Jan Jaap is acknowledged for his preliminary work in building, designing and testing the early generations of the intervalometer. Thanks must go to Disdro in Delft for use of their workshop and to Dirk and Shashwat for their help in building the newer version of the intervalometer that was used in this research. Special thanks to the hotels located on Mafia Island (Didimiza Guest House, Meremeta Lodge, Eco Shamba Kilole Lodge, Kinasi Lodge, Pole Pole Bungalows and the Mafia Island Lodge) and Chole Island (Chole Mjini Treehouse Lodge) for kindly allowing me access to their land and their support in setting up the experiment during the monsoon season. Yusufu, thank you for your help in making frames to support each of the instruments and digging holes with me in the hot sun. Thanks to Brenda and Jem for giving me a place to stay on Mafia Island. Thank you to Marco for having a soldering iron in the middle of nowhere and allowing me to use your tools so that I could resurrect some of the intervalometers that had been badly damaged during an intense storm.

In Delft, I would like to thank my thesis assessment committee of Nick, Marie-Claire and Marc for their guidance, input and patience. In particular, I would like to thank Nick for giving me the opportunity to conduct this research on the island I grew up on in Tanzania and also for providing me with interesting work through TAHMO so that I could support myself through the course of my studies. I would like to thank Lydia in the secretariat for helping me countless times and always with a smile. I would like to thank my friend Derron for the many meals where we talked through our respected thesis topics together. These conversations pushed me to think clearer and deeper. I would also like to thank Jasper, Julia and many others for making my life fun through all the hard work.

Finally, thanks go to my family for their love, support and belief in me. I never would have even dreamed of moving thousands of kilometers from home, with only enough money to survive for 6 months out of what has now become three years, to pursue this degree if it were not for knowing that you all had my back.



The work leading to these results has received funding from the European Community's Horizon 2020 Programme (2014-2020) under grant agreement n° 776691. The opinions expressed in the document are of the authors only and no way reflect the European Commission's opinions. The European Union is not liable for any use that may be made of the information.



TWIGA

Preface

The core scientific work of this master of science (MSc) thesis has been written up in the form of a journal paper. This paper is titled: *Is something fishy going on? Evaluating the Poisson hypothesis for rainfall estimation using intervalometers: first results from an experiment in Tanzania*. Hereafter referred to as "The paper". The paper will be submitted to the Atmospheric Measurement Techniques journal of the European Geosciences Union shortly after the conclusion of the MSc defense. The paper has been included in this thesis under chapter 1. Chapter 1 is the core work of this thesis and can be read as a stand-alone document. The remainder of this thesis is taken up by supplementary materials in the form of introductory notes and an appendix.

Appendix A contains a README for the python code that was developed during the course of the MSc research. Several thousand lines of code were written and, as much as possible, the author has attempted to write the code in such a way that it is easily legible to anybody who may be seeking to build on this research or re-analyse the data. This means that the code is well spaced and sensible variables names have been chosen and consistent naming hierarchy logic has been employed to the best of the author's ability. Despite these efforts it may still be difficult to understand the logical flow and purpose of each script or function individually as well as how they relate to one another. Therefore, it was deemed a necessary kindness to provide a README for the code in which the different tasks that each python script or function completes are explained as well as the overall logical flow in the data analysis that links one script or function to the next. The actual code files are submitted as further supplementary materials alongside this thesis and can be found in the education repository at <https://repository.tudelft.nl>, by searching for the title of this MSc thesis.

Contents

1	The Paper	1
1.1	Abstract	1
1.2	Introduction	1
1.3	Materials	3
1.3.1	Instruments	3
1.3.2	Experiment	3
1.3.3	Data Availability	4
1.4	Methods	5
1.4.1	Deriving rainfall rates from rain drop arrival rates	5
1.4.2	Calibrating the Intervalometer	6
1.4.3	Testing the Poisson homogeneity hypothesis	9
1.5	Results	12
1.5.1	Rainfall Rates	12
1.5.2	Testing the Poisson Hypothesis	15
1.6	Discussion	21
1.6.1	Rainfall Rates	21
1.6.2	Testing the Poisson Hypothesis	22
1.7	Conclusions	24
1.8	References	26
A	README for the python code files	28
A.1	Overview of the code	28
A.2	Description of each script and function	29
A.2.1	Script Name: import_DC.py	29
A.2.2	Script Name: import_disdro.py	29
A.2.3	Script Name: import_tb.py	30
A.2.4	Script Name: continuous_record_dc.py	30
A.2.5	Script Name: continuous_record_disdro.py	31
A.2.6	Script Name: stations_online.py	31
A.2.7	Script Name: Poisson_Testing.py	32
A.2.8	Script Name: PoissonAnalysis.py	32
A.2.9	Script Name: RainfallAnalysis.py	33
A.2.10	Function Name: seperate_rain_events.py	34
A.2.11	Function Name: poisson_test.py	34
A.2.12	Function Name: expon_rain.py	35
A.2.13	Function Name: exp_poly_constants.py	35
A.2.14	Function Name: exp_drop_size.py	36
A.2.15	Function Name: expon_rain_adj.py	36
A.2.16	Function Name: rain_adj_final.py	36

Is something fishy going on? Evaluating the Poisson hypothesis for rainfall estimation using intervalometers: first results from an experiment in Tanzania.

Didier de Villiers¹, Rolf Hut¹, Marc Schleiss², Marie-claire ten Veldhuis¹, and Nick van de Giesen¹

¹Department of Water Management, Faculty of Civil Engineering, TU Delft

²Department of Geoscience & Remote Sensing, Faculty of Civil Engineering, TU Delft

Correspondence: Didier de Villiers (devilliersdidier@gmail.com)

Abstract. Sub-Saharan Africa is one of the most vulnerable regions in the world to climate change. This is largely driven by the dependence on rain fed agriculture for food production. At the same time African climate observational networks have been in decline since the 1990s. A new kind of rainfall sensor (the intervalometer), which counts the arrival of drops at a piezo electric element, is tested during the Tanzanian monsoon season alongside tipping buckets and an impact disdrometer.

5 Rainfall rates are derived from rainfall arrival rates using Marshall and Palmer's (1948) exponential parameterisation. This parameterisation is defined independently of a notion of scale and therefore implicitly assumes that rainfall is a homogeneous Poisson process. Testing of the Poisson assumption shows that 22.5% of the total drops observed can reasonably be considered Poisson and that the main reason for Poisson deviations are non compliance with the stationarity criterion (36.7%) and the presence of correlations between drop counts (14.3%), particularly at higher arrival rates ($\rho_a > 500 [m^{-2}.s^{-1}]$). The total
10 rainfall amount [mm] calculated from intervalometer measurements overestimates the tipping bucket value by a factor of approximately three. The overestimate is most likely due to poor calibration of the minimum detectable drop size (D_{min}). A correction is applied to constrain the overestimates of mean drop size by the intervalometer parameterisation to the observed disdrometer measurements. The correction results in an improvement in the estimate of the total rainfall amount to within 10% of tipping bucket measurements. The total rainfall amount [mm] calculated from disdrometer rainfall arrival rates is within 5%
15 of co-located tipping bucket measurements. The form of the mean drop size relation with arrival rate appears stable in time and space. The intervalometer shows good potential for use as a rainfall measurement instrument or to derive estimates of mean drop sizes.

1 Introduction

Africa and particularly Sub-Saharan Africa is one of the most vulnerable regions in the world to climate change (Boko et al.,
20 2007). This is predominantly because the main economic activity (by share of labour) is rainfed agriculture-98% of crop production is rainfed (Abdrabo et al., 2014). At the same time much of Sub-Saharan Africa is greatly under serviced by climate observations and the existing observational networks have been in decline since the mid 1990s; from 8 stations reporting per 1,000,000 km² to 1 or even none in some grid boxes in 2015 (data from the Climate Research Unit of the University of East

Anglia, 2017). There are some organisations working on setting up new observational networks, such as the Trans-African
25 Hydro-Meteorological Observatory (TAHMO) but progress is slow due to the lack of existing value chains for weather data
(TAH, 2017). In general, African climate has not been well researched (Otto et al., 2015; Washington et al., 2006).

There is a need for robust, inexpensive and accurate rainfall measuring instruments. For example, a recent review into the
scaling up of index insurance for smallholder farmers (some of the world's poorest people) found that the sparsity of ground
based weather stations is a large challenge for insurers in Sub-Saharan Africa (Greatrex et al., 2015) and companies have been
30 forced to look to other sources of data or to develop other indices by which to insure crops. Satellite missions, such as the
Global Precipitation Measurement (GPM) mission show good potential for bridging this gap. However, satellite observations,
whilst providing good spatial coverage, do not cover the entire temporal period and the spatial resolution may be too coarse for
some applications.

Satellite data faces another issue for areas with a lack of ground based data for validation. Since radars do not measure
35 rainfall directly, rainfall estimates are dependent on an accurate parameterisation of the drop size distribution (DSD) in order
to develop rainfall (R) to radar reflectivity (Z) relationships (Munchak and Tokay, 2008; Guyot et al., 2019). A foundational
work in this regard is the negative exponential parameterisation presented by Marshall and Palmer 1948 as a fit to filter paper
measurements of rain drop sizes for different rain rates. A lot of work has been done on determining the functional forms
for these parameterisations and many different forms of the DSD have been proposed, of which the most widely used are the
40 aforementioned exponential, gamma (Ulbrich, 1983; Tokay and Short, 1996; Iguchi et al., 2017) and lognormal distributions
(Feingold and Levin, 1986). It has also been shown that the appropriate parameterisation is dependent on the type of rainfall
(Atlas and Ulbrich, 1977) and the climatic setting (Battan, 1973; Bringi et al., 2003). Therefore, ground 'truthing' of DSDs for
satellite retrievals is very important to ensure that the DSD is being parameterised correctly in the derivation of rainfall rates
(Munchak and Tokay, 2008).

45 An assumption that is seldom explicitly mentioned in the presentation of these parameterisations is the homogeneity as-
sumption (Uijlenhoet et al., 1999). The concept of the DSD is only useful if at some minimum scale raindrops are distributed
homogeneously in space and time. If this was not the case then the parameterisation would depend on the size of the sample
volume/area/time period to which it pertains (Uijlenhoet et al., 1999). Statistical homogeneity implies that the frequency of
raindrops in a volume or arriving at a surface in fixed time intervals obeys Poisson statistics. The arrival of raindrops at a
50 surface has long been considered an example of a Poisson process (Kostinski and Jameson, 1997; Joss and Waldvogel, 1969).
However, this assumption has been questioned and several studies argue that the homogeneity assumption is unable to cope
with the clumping of raindrops both in time and space that is observed in reality. To borrow Jameson and Kostinski's (1997) ex-
ample; the 'streakiness' that is part of the lived experience of rainfall can be seen when sheets of rain pass across the pavement
during thunderstorms. This clumping results in greater variability than is expected by Poisson statistics.

55 Generally two different approaches have been taken to explain the enhanced variability. Studies like (Lovejoy and Schertzer,
1990; Lavergnat and Golé, 1998) propose to abandon the Poisson process framework and replace it with a scale dependent,
multi-fractal framework. And others, that propose to generalize the homogeneous Poisson process (with a constant mean)
to a doubly stochastic Poisson process or Cox process, where the mean itself is a random variable (Jameson and Kostinski,

1998; Smith, 1993). The implications for radar meteorology of abandoning the Poisson framework would require an entire re-working of how rainfall estimates are derived.

The aim of this study is to formally assess the adequacy of the Poisson assumption and its importance in deriving rainfall estimates from ground based measurements. To this end nine intervalometers were deployed over a two month period during the Tanzanian tropical monsoon.

2 Materials

2.1 Instruments

Three different types of instrument were used in the experiment; a Tipping Bucket rain gauge made by Onset, in the US, equipped with a HOBO datalogger; an Acoustic Disdrometer made by Disdro in Delft, The Netherlands; and an Intervalometer. The intervalometer is a simple device that registers the arrival of raindrops at the surface of a piezo electric drum. It has a minimum detectable drop diameter of 0.8 mm, determined by Pape (2018) in a lab experiment. Typical values of D_{min} for impact disdrometers are between 0.3 and 0.6mm (Johnson et al., 2011). The larger than typical D_{min} value for the intervalometer means that the instrument may give slight underestimates of long term rainfall rates. The advantage of the intervalometer over a standard rain gauge, is that the drop counts can be used to constrain radar observations. Furthermore, the combination of intervalometer measurements with rain gauge data can be used to give crude estimates of the observed mean drop sizes. More information about the intervalometer can be found at <https://github.com/nvandegiesen/Intervalometer/wiki/Intervalometer> or in Pape's (2018) report. The acoustic disdrometer, registers the kinetic energy of drop impacts at a drum and converts this to an estimate of the drop size. It can be thought of as a intervalometer that not only counts drops but also provides estimates of the drop size. The minimum detectable drop diameter for the disdrometer is 0.6mm. The tipping bucket rain gauge collects all drops over a surface area and funnels it to a small bucket with a resolution of 0.2mm. When the bucket is full, it tips over. The volume of each tip is verified in situ via a field calibration experiment (WMO, 2014). A good discussion of the pros and cons of impact disdrometers can be found at e.g. (Tokay et al., 2001; Guyot et al., 2019) and for tipping buckets at e.g. (Sevruk, 1985; WMO, 2014). In total, the experiment made use of nine intervalometers, one acoustic disdrometer and two tipping bucket rain gauges at eight different sites.

2.2 Experiment

Eight sites were selected along the southern coast of Mafia Island, Tanzania in an approximate line, such that a rectangle 3.1 km in length and 500m in width would cover all the sites. The dimension of the long axis of the experiment was chosen to be approximately that of the spatial resolution ($5km$) of GPM mission's dual polarization radar (DPR) instrument.

Table 1. The rainfall measurement sites

<i>Site Name and Coordinates</i>	<i>Instruments</i>	<i>Distance [m] from nearest site</i>
Didimiza [$-7.9792S, 39.7317E$]	Intervalometer	Meremeta [421m]
Meremeta [$-7.9756S, 39.7327E$]	Intervalometer	Didimiza [421m]
Shamba Kilole [$-7.9757S, 39.7383E$]	Intervalometer	Meremeta [621m]
Kinasi [$-7.9767S, 39.7444E$]	Intervalometer	Pole Pole [90m]
Pole Pole [$-7.9767S, 39.7452E$] (Main Site)	Tipping Bucket, Disdrometer and Intervalometer	Kinasi [90m]
MIL1 [$-7.9732S, 39.7485E$]	Tipping Bucket and 2 Intervalometers	MIL2 [72m]
MIL2 [$-7.9726S, 39.7487E$]	Intervalometer	MIL1 [72m]
Chole Mjini [$-7.9718S, 39.7584E$]	Intervalometer	MIL2 [1070m]

2.2.1 Site Selection

Rainfall measurement sites were chosen to comply as much as possible with World Meteorological Organisation guidelines within the constraints of accessibility and landscape. Ideally, this means that all of the sensors should be placed in clearings, sheltered as much as possible from the wind at a height of 1.5m off the ground and 1.5m to the nearest instrument (if co-located) and between $2 \times H$ and $4 \times H$ from the nearest object, where H is the height of the nearest obstacle above the surface of the rainfall measurement instrument (WMO, 2014). All guidelines were complied with except for the H requirement due to dense vegetation within the entire observational area. In practice, the distance to the nearest object ranged between H and $2 \times H$. No instruments were placed at sites where the nearest obstacle was $\leq H$ away. Tipping buckets were calibrated in the field.

2.3 Data Availability

There were some issues over the course of the experiment with the various instruments which affect the availability of data. The disdrometer picked up on an oscillating signal from the 20/05/2018 onwards that resulted in total corruption of the data. Some intervalometers experienced water damage, particularly in storms with high rainfall intensities, which caused the instruments to go offline for certain periods and two were damaged beyond repair. The Tipping Bucket gauges experienced no known issues. The complete data record is presented in figure 1.

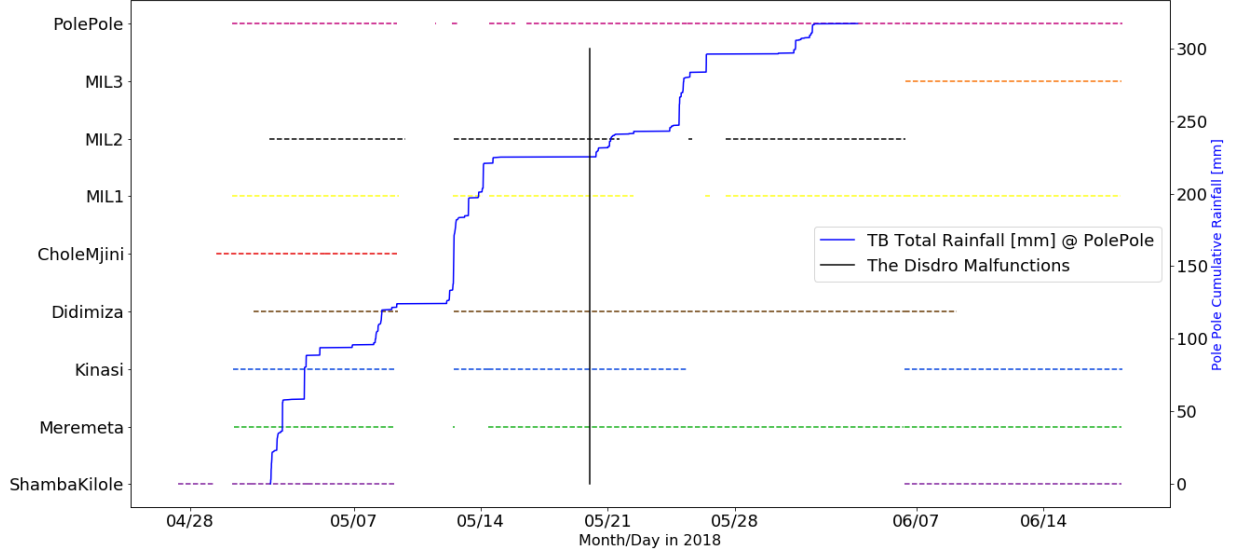


Figure 1. Record of "Instrument Online" for each intervalometer site and the total rainfall amount [mm] from the tipping bucket at Pole Pole.

3 Methods

3.1 Deriving rainfall rates from rain drop arrival rates

Uijlenhoet and Stricker (1999) present an excellent review of the exponential parameterization of the DSD as well as full
 105 derivations in their paper. A small summary mostly derived from their work is presented below. The raindrop size distribution
 in a volume of air $N_V(D)[mm^{-1}.m^{-3}]$ is defined such that the quantity $N_V(D)dD$ represents the number of drops with diam-
 eters between D and dD per unit volume of air. Marshall and Palmer (1948) proposed a negative exponential parameterisation
 for $N_V(D)$, based on filter paper measurements, of the form:

$$N_A(D) = N_0 \exp(-\Lambda D), \text{ where} \quad (1)$$

$$110 \quad \Lambda = 4.1 R^{-0.21} [mm^{-1}] \quad (2)$$

$$N_0 = 8 \times 10^3 [mm^{-1}m^{-3}] \quad (3)$$

If raindrops are assumed to fall at terminal velocity then $N_V(D)$ can be related to the DSD of drops arriving at a surface
 $N_A(D)[m^{-2}.s^{-1}]$ by $v(D)$, which describes the relationship between drop diameter and terminal fall velocity. $N_A(D)$ is the
 form of the DSD that is observed by disdrometers and intervalometers (Uijlenhoet and Stricker, 1999; Smith, 1993).

$$115 \quad N_A(D) = v(D)N_V(D) \quad (4)$$

$$v(D) = \alpha D^\beta \quad (5)$$

Atlas and Ulbrich (1977) showed that $\alpha = 3.778[m.s^{-1}mm^{-\beta}]$ and $\beta = 0.67[-]$ provide a close fit to the data of Gunn and Kinzer (1949) for $0.5mm \geq D \geq 5.0mm$. The mean rainfall arrival rate $\rho_A[m^{-2}.s^{-1}]$ is defined as the integral over all drop sizes of $N_A(D)$. For the intervalometer this is the integral between $D_{min} = 0.8$ and ∞ since the instrument has a minimum detectable drop diameter.

$$\rho_A = \int_{D_{min}}^{\infty} N_A(D)dD \quad (6)$$

$$\rho_A = \alpha N_0 \int_{D_{min}}^{\infty} D^\beta \exp(-\Lambda D)dD \quad (7)$$

$$\rho_A = \alpha N_0 \frac{\Gamma(1 + \beta, \Lambda D_{min})}{\Lambda^{1+\beta}} \quad (8)$$

Where Γ is the upper incomplete gamma function. Uijlenhoet and Stricker (1999) showed that for self consistency purposes, the use of $= 4.1R^{-0.21}$ determines that $\alpha = 3.25, \beta = 0.762$, which are quite similar to the values presented by Atlas and Ulbrich (1977). Using the Uijlenhoet and Stricker (1999) α, β values and the Marshall and Palmer (1948) $R - \Lambda$ relationship the rainfall rate (R) can then be calculated from the measured rainfall arrival rate ρ_A by first calculating the rainfall arrival rate at different rainfall rates ($0 - 200mm/h$ in a step of $0.01mm/h$) and then fitting a third order polynomial to the curve in Python using polyfit. The fit is forced through the origin and returns a correlation coefficient of approximately 1. The constants of the fitted polynomial can then be used to calculate the rainfall rate from the measured arrival rate. A plot of the $\rho_A - R$ curve and fitted polynomial is shown in figure 2.

3.2 Calibrating the intervalometer

The intervalometer is still in development as an instrument and therefore, if it is giving poor estimates, then a more trusted and proven instrument can be used for calibration. Sources of measurement error for the intervalometer are the calibration of the parameter D_{min} and the measurement of ρ_A . Errors in the determination of D_{min} affect the $\rho_A - R$ relationship. Errors in the rainfall arrival rate can result from, splashing of drops from outside the sensor onto the sensor surface during high intensity rainfall (resulting in overestimates), spurious drops from something other than rain falling on the sensor (resulting in overestimates), or from edge effects (resulting in underestimates). Drops with $D > D_{min}$ landing near the edges of the sensor have a dampened signal and may not be recorded if D is quite close to D_{min} . If there are *a priori* disdrometer measurements of the mean drop size for different rainfall intensity or arrival rates these can be used to constrain the intervalometer measurements and provide more accurate rainfall estimates. The observed mean drop sizes can be incorporated into the parameterisation to ensure that the expected mean drop sizes of the parameterised gamma distribution match with the disdrometer measurements.

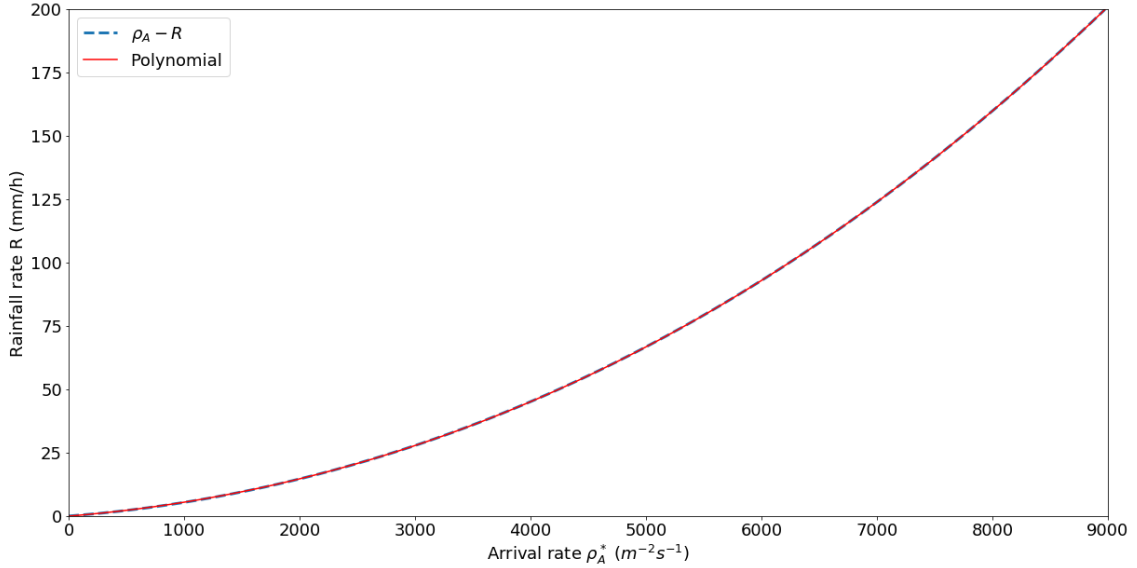


Figure 2. The third order polynomial fitted to equation 8, used to calculate rainfall rate [mm/hr] from rainfall arrival rate [$m^{-2}.s^{-1}$]. The correlation co-efficient has a value of 1, rounded to the sixth decimal place.

At a measured ρ_A the expected (mean) drop size can be calculated by recognising that the probability distribution of the stochastic drop diameters arriving at a surface per unit time $f_D(A) = \rho_A^{-1} N_A(D)$ is a gamma distribution (Uijlenhoet and Stricker, 1999); in this case truncated at D_{min} .

$$f_D(A) = \frac{\Lambda^{1+\beta}}{\Gamma(1+\beta, \Lambda D_{min})} \times D^\beta \exp(-\Lambda D) \quad (9)$$

$$\beta, \Lambda > 0, D \geq D_{min} \quad (10)$$

The expected value (mean) of a left truncated gamma distribution and complete gamma distribution is given by e.g. (Johnson et al., 2011; Uijlenhoet and Stricker, 1999):

$$\mu_{D_{A,exp}} = E[D_{A,exp} > D_{min}] = \left(\frac{1+\beta}{\Lambda} \right) \frac{1 - \frac{\gamma(2+\beta, \Lambda D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda D_{min})}{\Gamma(1+\beta)}} \quad (11)$$

$$\mu_{D_{A,exp}} = E[D_{A,exp}] = \frac{1+\beta}{\Lambda} \quad (12)$$

Where γ is the lower incomplete gamma function. Now, if the observed mean drop sizes ($\mu_{D_{A,obs}}$) are some function of $\rho_{A,obs}$, $f(\rho_{A,obs})$ then we can express the expected rainfall rate (R_{exp}) and a 'corrected' rainfall rate (R_{corr}) as functions of the

expected and observed mean drop sizes by using the relationship $\Lambda = 4.1R^{-0.21}$. A good first guess for the form of $f(\rho_{A,obs})$ is the expectation of the gamma parameterisation above, but could be any function or simply the observed data at each rainfall arrival rate. For the complete gamma distribution an analytical solution exists.

$$R_{exp} = \left(\frac{1+\beta}{\mu_{D_A,exp}} \times \frac{1}{4.1} \right)^{\frac{-1}{0.21}} \quad (13)$$

$$R_{corr} = \left(\frac{1+\beta}{\mu_{D_A,obs}} \times \frac{1}{4.1} \right)^{\frac{-1}{0.21}} \quad (14)$$

Divide R_{corr} by R_{exp} to get:

$$\frac{R_{corr}}{R_{exp}} = \frac{\left(\frac{1+\beta}{\mu_{D_A,obs}} \times \frac{1}{4.1} \right)^{\frac{-1}{0.21}}}{\left(\frac{1+\beta}{\mu_{D_A,exp}} \times \frac{1}{4.1} \right)^{\frac{-1}{0.21}}} \quad (15)$$

$$\frac{R_{corr}}{R_{exp}} = \left(\frac{\mu_{D_A,exp}}{\mu_{D_A,obs}} \right)^{\frac{-1}{0.21}} \quad (16)$$

$$R_{corr} = R_{exp} * D_{eff} \quad (17)$$

$$D_{eff} = \left(\frac{\mu_{D_A,exp}}{\mu_{D_A,obs}} \right)^{\frac{-1}{0.21}} \quad (18)$$

$$\mu_{D_A,obs} > \mu_{D_A,exp} \rightarrow D_{eff} > 1 \quad (19)$$

$$\mu_{D_A,obs} < \mu_{D_A,exp} \rightarrow 0 \leq D_{eff} \leq 1 \quad (20)$$

D_{eff} is an effective parameter that scales the expected rainfall rates (calculated from the parameterisation of the DSD) to the observed DSD. Alternatively, if the intervalometer is co-located with a rain gauge then the independent observations of rainfall can be used to give an estimate of the mean drop size relation by re-arranging equation 16.

$$\mu_{D_A,exp} = \mu_{D_A,obs} \times (R_{eff})^{-0.21} \quad (21)$$

$$R_{eff} = \left(\frac{R_{obs}}{R_{exp}} \right)^{-0.21} \quad (22)$$

There is a small time delay between the first registering of drops on the surface of the intervalometer and the first tip of the tipping bucket due to the small volume of the bucket. Therefore it is not recommended to directly compare instantaneous rainfall rates (Ciach, 2003). However, by averaging over longer time periods such as the length an entire day, reasonable total rainfall amounts can be obtained in order to calculate R_{eff} in equation 22.

For the truncated gamma distribution a numerical approach is required and can be implemented in Python.

$$\mu_{D_{A,exp}} = \left(\frac{1+\beta}{\Lambda_{exp}} \right) \frac{1 - \frac{\gamma(2+\beta, \Lambda_{exp} D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda_{exp} D_{min})}{\Gamma(1+\beta)}} \quad (23)$$

$$R_{exp} = \left(\left(\frac{1+\beta}{\mu_{D_{A,exp}} \times 4.1} \right) \frac{1 - \frac{\gamma(2+\beta, \Lambda_{exp} D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda_{exp} D_{min})}{\Gamma(1+\beta)}} \right)^{-\frac{1}{0.21}} \quad (24)$$

$$R_{corr} = \left(\left(\frac{1+\beta}{\mu_{D_{A,obs}} \times 4.1} \right) \frac{1 - \frac{\gamma(2+\beta, \Lambda_{corr} D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda_{corr} D_{min})}{\Gamma(1+\beta)}} \right)^{-\frac{1}{0.21}} \quad (25)$$

Divide R_{corr} by R_{exp} to get:

$$\frac{R_{corr}}{R_{exp}} = \left(\frac{\left[\frac{1}{\mu_{D_{A,obs}}} \frac{1 - \frac{\gamma(2+\beta, \Lambda_{corr} D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda_{corr} D_{min})}{\Gamma(1+\beta)}} \right]}{\left[\frac{1}{\mu_{D_{A,exp}}} \frac{1 - \frac{\gamma(2+\beta, \Lambda_{exp} D_{min})}{\Gamma(2+\beta)}}{1 - \frac{\gamma(1+\beta, \Lambda_{exp} D_{min})}{\Gamma(1+\beta)}} \right]} \right)^{-\frac{1}{0.21}} \quad (26)$$

$$\Lambda_{corr} = 4.1 \times R_{corr}^{-0.21} \quad (27)$$

Note that equation 26 is the same as equation 16 with an extra term to account for the truncation. The D_{min} values in the above equation are 0.6mm for the observed drop sizes (from the disdrometer) and 0.8mm for the expected drop sizes (from the intervalometer). R_{exp} , Λ_{exp} and $\mu_{D_{A,exp}}$ are calculated from the observed rainfall arrival rate and relevant equations. Then if $\mu_{D_{A,exp}} < \mu_{D_{A,obs}}$ guess $R_{corr} = R_{exp}$ and calculate Λ_{corr} and the right hand side (RHS) of equation 26. Iterate by increasing R_{corr} by 0.01 [mm/h] until the LHS and RHS are equal. if $\mu_{D_{A,exp}} > \mu_{D_{A,obs}}$ then guess $R_{corr} = R_{exp}$ and calculate Λ_{corr} and the RHS of equation 26. Iterate by decreasing R_{corr} by 0.01 [mm/h] until the LHS and RHS are equal. The final value of R_{corr} is the 'corrected' rainfall rate. Equation 26 can also be used to derive estimates of the 'corrected' mean drop size by incorporating co-located tipping bucket measurements R_{obs} with intervalometer estimates R_{exp} . Note, since the tipping bucket measures all drops it has no minimum detectable drop size and therefore we can combine equation 11 and equation 12 with the $R - \Lambda$ relationship to get:

$$\mu_{D_{A,corr}} = \frac{\mu_{D_{A,exp}} \times \left[1 - \frac{\gamma(1+\beta, \Lambda_{exp} D_{min})}{\Gamma(1+\beta)} \right]}{\left(\frac{R_{obs}}{R_{exp}} \right)^{-0.21} \times \left[1 - \frac{\gamma(2+\beta, \Lambda_{exp} D_{min})}{\Gamma(2+\beta)} \right]} \quad (28)$$

$$\Lambda_{exp} = 4.1 \times R_{exp}^{-0.21} \quad (29)$$

The 'corrected' drop sizes can be calculated directly from the above equation.

3.3 Testing the Poisson homogeneity hypothesis

The concept of a drop size distribution depends on the assumption that at some minimum spatial or temporal scale (the primary element) the data is homogeneous. Homogeneity in a statistical sense implies that the data within the primary element

follows Poisson statistics (Uijlenhoet and Stricker, 1999). In order for a process to be reasonably assumed as Poisson some key assumptions must hold. As applied to rainfall, these are as follows:

- 200 1. The random process is stationary
2. The event counts in non-overlapping time intervals are statistically independent
3. The probability of an event occurring during a time interval $t, t + \delta t$ is proportional to δt
4. The probability of more than one event in a time interval $t, t + \delta t$ becomes negligible for sufficiently small δt

If these fundamental assumptions hold then the distribution of event counts (rain drops) is given by (eg. (Feller, 2010)).

$$205 \quad p(k) = \frac{\mu^k \exp(-\mu)}{k!} \quad (30)$$

Where μ is the mean value per unit time and k is random number of drops observed during a particular counting interval/window of time. Kostinski and Jameson (1997) show that this evenly mixed Poisson model does not explain the clumpiness and super-Poisson variability that is observed in real rainfall. However, if μ itself is an unpredictable, random variable in time and space then a rainfall event can be sub-divided into N patches, each with its own μ . In order to derive an unconditioned PDF of the drop counts it is necessary to integrate over the probability distribution of the patches $f(\mu)$.

$$210 \quad p(k) = \int_0^{\infty} \frac{\mu^k \exp(-\mu)}{k!} f(\mu) d\mu \quad (31)$$

The variance of the Poisson mixture is enhanced beyond the variance of a pure Poisson PDF. Kostinski and Jameson (1997) show that the Poisson mixture provides a better description of the frequency of drop arrivals per unit time than a simple Poisson model. The definition of $f(\mu)$ in equation 31 implies that there is a definable coherence time τ over which μ can be considered stationary and to which the simple Poisson model can be applied. In order to estimate $f(\mu)$ with sufficient accuracy require $(t \ll \tau \ll T)$. Where T is the entire length of a rainfall event, τ is the coherence time of a patch and t is the counting interval for the raindrops. Kostinski and Jameson (1997) show that an order of magnitude difference is sufficient between t, τ and τ, T . For the intervalometer data, rain drops are aggregated into 10 second bins. Therefore, the minimum value for τ is 100s and for T it is 1000s. The length of τ can be determined by calculating the normalized auto-correlation function of a rainfall event of length T at increasing lag times. The lag time for which the auto-correlation drops below $\frac{1}{e}$ is defined as τ (Kostinski and Jameson, 1997). A rainfall event can then be sub-divided into N patches of length τ and the fundamental Poisson assumptions can be tested on each patch.

Assumptions 3,4 are trivial for rainfall and 1,2 can be tested. A hierarchical test is used, where a patch of rainfall, of length τ must pass each test before moving onto the next test. Note that since it is impossible to know where such a patch of length τ may start or end in the data record then it is best to view τ as a moving window over which the statistical tests are conducted. Upon conclusion of the test, the window shifts one data point forward in time and the tests are conducted again. This methodology also ensures that the number of effective samples is increased. The system of hierarchical tests is as follows.

1. Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests for stationarity, with a p-value of 0.05. Both test results must indicate stationarity.

230

- (a) The KPSS null hypothesis is that the process is trend stationary.
- (b) The ADF null hypothesis is that the series has a unit root (not stationary).

2. Auto-correlation function at increasing lag times must be within the 95% confidence limit (CL) of a Poisson process with n samples

235

(a)

$$\mu_{auto-corr} = \frac{-1}{n-1} \quad (32)$$

$$\sigma_{auto-corr}^2 = \frac{n-2}{(n-1)^2} \quad (33)$$

3. χ^2 test for goodness of fit between the observed frequencies and the expected frequencies of a Poisson distribution with the same mean, p-value = 0.05.

240

4. Dispersion criterion, such that the observed dispersion must be within the 95% CL of a Poisson distribution of n samples

(a)

$$\mu_{disp} = 1 \quad (34)$$

$$\sigma_{disp} = \left[\frac{2}{n-1} \right]^{\frac{1}{2}} \quad (35)$$

5. Calculation of Kullback-Leibler (KL) divergence to give a sample independent indication of how well the observed distribution matches the Poisson distribution. The KL divergence, also known as the relative entropy, between two probability density functions is commonly used as a measure of similarity or 'distance' between the distributions (Hershey and Olsen, 2007).

245

Tests 1 and 2 test the stationarity and independence assumptions of a Poisson process. Test 3 checks that the distribution matches a Poisson distribution and Tests 4 and 5 are quality checks to ensure that the tests are providing good results. The quality check is used because often the sample size over which each test is conducted is quite small. Figure 3 shows an example of a patch of rainfall that passes all of the tests and can therefore reasonably be assumed to comply with the Poisson Homogeneity assumption.

250

The rainfall can be characterised by uncorrelated fluctuations around a constant mean rate of arrival, in this case 365.7 $[m^{-2}.s^{-1}]$. The corresponding probability mass function (pmf) of this patch of rainfall along with the expected pmf function

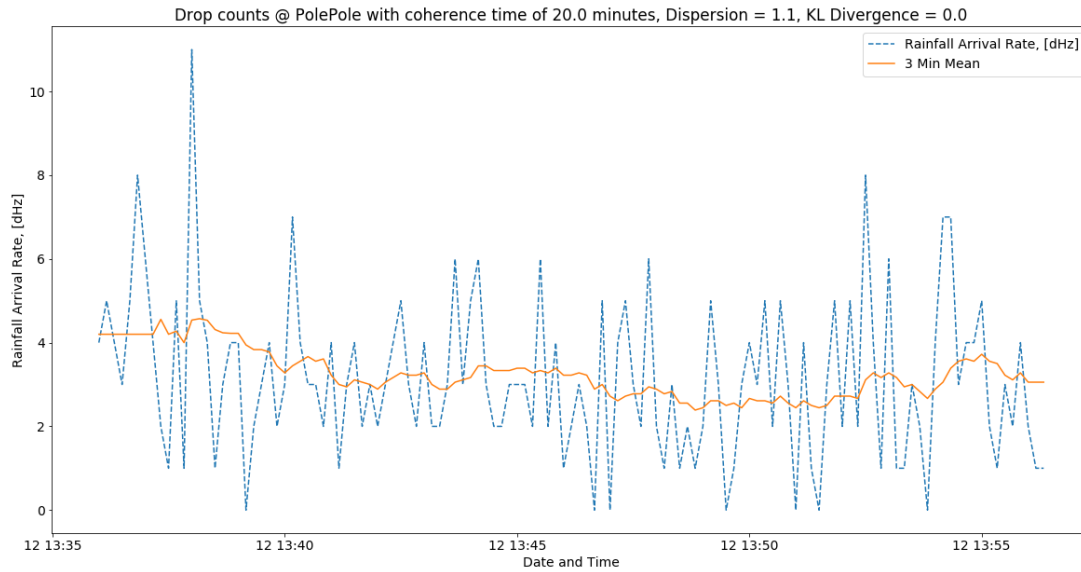


Figure 3. Rainfall arrival rate (10 second drop counts in blue) of a patch of rainfall, with a coherence time of 20 minutes, that can reasonably be assumed to be Poisson. The 3 min rolling mean is also plotted in the figure (in orange). The dispersion of the patch is 1.1 (expected value of 1) and the KL-divergence is 0.0, indicating very good agreement between the observed probability densities of the patch and the expected probability density from Poisson.

255 of a Poisson process with the same mean arrival rate is plotted in figure 4. These two figures are shown as an example of what
the patches of rainfall that pass all of the hierarchical tests look like.

4 Results

4.1 Rainfall Rates

The total rainfall amounts [mm] observed by the co-located tipping bucket, intervalometer and disdrometer at the main site
260 (Pole Pole) for the longest 'online' period of the three instruments are presented in figure 5. Estimates of total rainfall derived
from the disdrometer arrival rates are in good agreement with the tipping bucket record (the records match to within std error).
This is not the case for the intervalometer, which provides a large over-estimate of the total rainfall compared to the tipping
bucket (by a factor of almost 3). The figure also shows that the intervalometer registers much higher arrival rates than the
disdrometer over all rainfall events despite having a smaller sensor area and a larger minimum detectable drop size. Calibration
265 of the intervalometer rainfall estimate (using equation 26) by the observed mean drop sizes results in good agreement with
the tipping bucket record as a whole (within 8% of the tipping bucket value). In figure 6 the performance of the calibrated

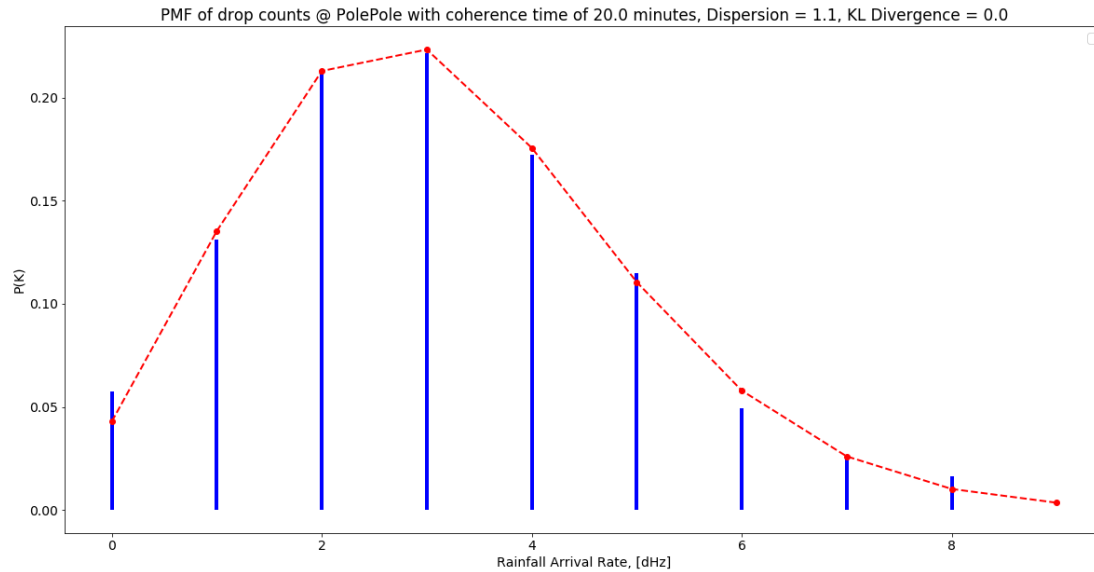


Figure 4. The Poisson Probability Mass Function (pmf in red) is plotted against the observed probability densities (in blue) of drop counts for a patch of rainfall that passes all the hierarchical tests at Pole Pole. The coherence time is 20 minutes, dispersion = 1.1, KL divergence = 0.0.

parameterisation over certain rainfall events with the observational record for Pole Pole (left side) and MIL1 (right side) is presented. For Pole Pole, the calibrated parameterisation also provides reasonable estimates over each rainfall event, except for in panel three where it significantly underestimates the tipping bucket value. The rainfall event in panel three is characterised by much lower arrival rates than the events in panels one and two.

The intervalometer calibration derived from the main site can be applied to the MIL1 site measurements (right hand side of fig 6), where tipping bucket measurements are also available. The calibration gives good estimates at this site over the record as a whole and over the different rainfall events within the time period (right side of fig 6). Note that MIL1 is situated approximately 1km from Pole Pole, where the mean drop sizes were observed by the disdrometer.

The total cumulative rainfall estimates over the entire online record of the intervalometers at Pole Pole and MIL1 are presented in table 2.

The table shows that the calibrated intervalometer results are in good agreement with the tipping bucket values. Within 5% for Pole Pole and within 9% for MIL1.

The intervalometer estimates and co-located tipping bucket measurements are also used to derive estimates of the mean drop sizes, via equation 28. The estimated drop sizes and the corresponding line of best fit (using the form of equation 11) are plotted alongside the observed mean drop sizes as well as the expected values from the parameterisation in figure 7. The drop size

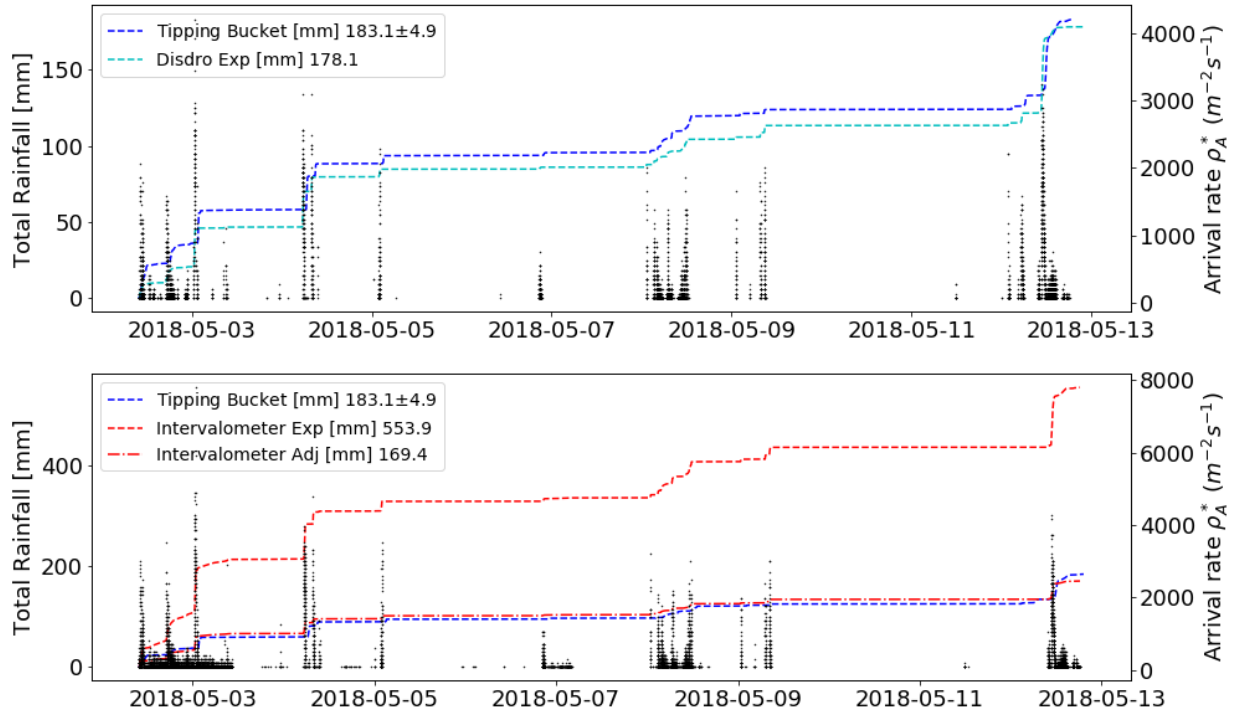


Figure 5. The total rainfall amounts [mm] observed by the co-located tipping bucket, intervalometer and disdrometer at the main site (Pole Pole) for the longest 'online' period of the three instruments are presented. The top panel shows the tipping bucket record against the disdrometer record and the bottom panel shows the uncalibrated and calibrated intervalometer record against the tipping bucket. Also, plotted (in black) are the rainfall arrival rates measured by the disdrometer and intervalometer, respectively. The suffixes 'Exp' and 'Adj' refer to the un-calibrated rainfall from the exponential parameterisation and the calibrated rainfall, respectively.

Table 2. Total cumulative rainfall [mm] over the online period of the intervalometers compared to the tipping bucket at MIL1 and Pole Pole

<i>Instrument and Parameterisation</i>	<i>Pole Pole (Main Site)</i>	<i>MIL1</i>
Intervalometer, Uncalibrated	802.0	144.6
Intervalometer, Calibrated	253.5	49.6
Tipping Bucket	266.7 ± 7.1	45.3 ± 1.3

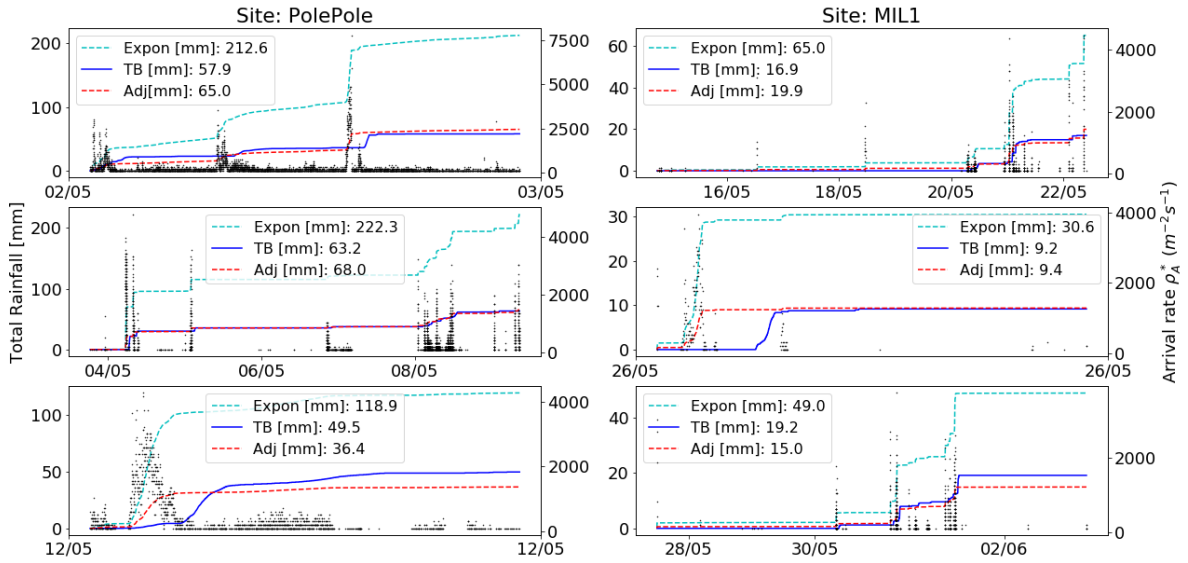


Figure 6. The calibration of the exponential parameterisation (labelled as Adj) is used to estimate rainfall rates for three different rainfall events at Pole Pole (left three panels) and MIL1 (right three panels). The calibrated rainfall is compared to the tipping bucket (TB) measurements as well as the un-calibrated measurements (Expon). Once again, the rainfall arrival rates are also plotted.

estimates are derived from only 22 data points. This is because it is necessary to aggregate tipping bucket measurements into daily averages in order to compare between the tipping bucket and intervalometer measurements. The estimated drop sizes show a large spread, particularly at arrival rates less than $500 [m^{-2}.s^{-1}]$. Above this arrival rate the estimated data shows reasonable agreement with the observed values. The best fit estimator gives $\beta = 0.37, D_{min} = 0.53$ for the intervalometer.

4.2 Testing the Poisson Hypothesis

The coherence time or window length over which the Poisson tests were performed ranged from 2-22 minutes across the sites, with a typical length being in the order of 10 minutes. The rainfall that passes all the tests and can therefore reasonably be assumed belong to a Poisson process has a mean dispersion value of 1.01 (expected value = 1) and a mean Kullback-Leibler divergence of 0.02 (expected value of 0). This indicates that the tests have resulted in patches of rainfall that are stationary, exhibit no correlation between drop counts within a 95% confidence interval, match a Poisson distribution very well and have a mean dispersion of approximately 1. The proportion of raindrops, averaged across all the intervalometers, that fail each of the hierarchical tests for 'Poisson-ness' as well as the mean arrival rate for each group is presented in figure 8. Overall, only 22.5% of all the observed rainfall can reasonably be assumed to be Poisson. 36.7% of rainfall patches fail because they are not stationary and 14.3 % do not pass the independence criterion, indicating the presence of correlations between drop counts. In

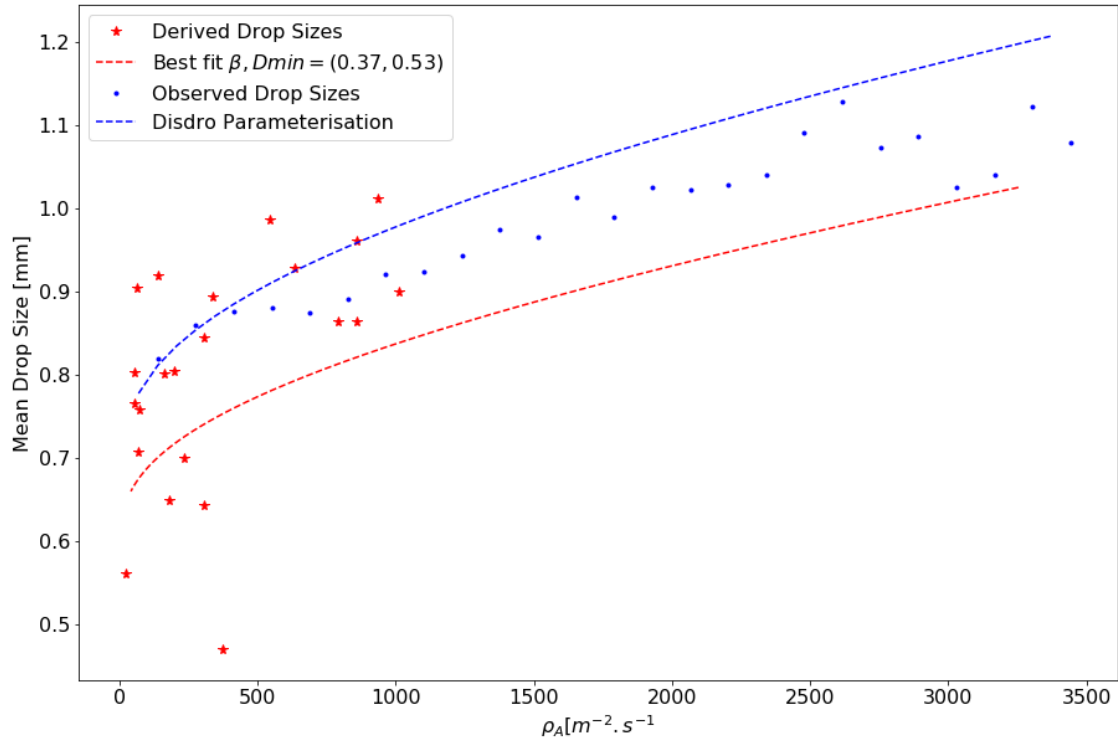


Figure 7. Estimates of the mean drop size against rainfall arrival rate calculated from equation 28, using a combination of intervalometer estimates and tipping bucket measurements of total rainfall amount are presented. The derived estimates and the line of best fit are plotted in red. Also shown are the measured and expected mean drop sizes from the disdrometer and exponential parameterisation, respectively. These are plotted in blue.

total 51% of all rainfall patches fail the tests due to the changes in the mean arrival rate or the presence of correlations between drop counts on scales as small as 2 minutes. It should be noted that these patches of rainfall are characterised by higher arrival rates (e.g. the rainfall that fails the independence test has a mean ρ_A that is approximately 3 times the Poisson value). Of the remaining 49 % of the rainfall patches approximately half do not fit a Poisson distribution, and the other half are classified as Poisson. A very small subset (3.1 %) do not pass the dispersion criteria due to being over-disperse. I.e. the observed variance is larger than what is expected for Poisson statistics. Again, this rainfall is characterised by higher rainfall arrival rates.

These average values are quite representative for all the sites, except for Chole Mjini. This site is atypical in that it was only online for a relatively short period between the 30/04/2018 and the 08/05/2018 and during this period 77% of all the rainfall was classified as Poisson. This can be clearly seen in the two middle panels of figure 9. The time series of rainfall arrival rates

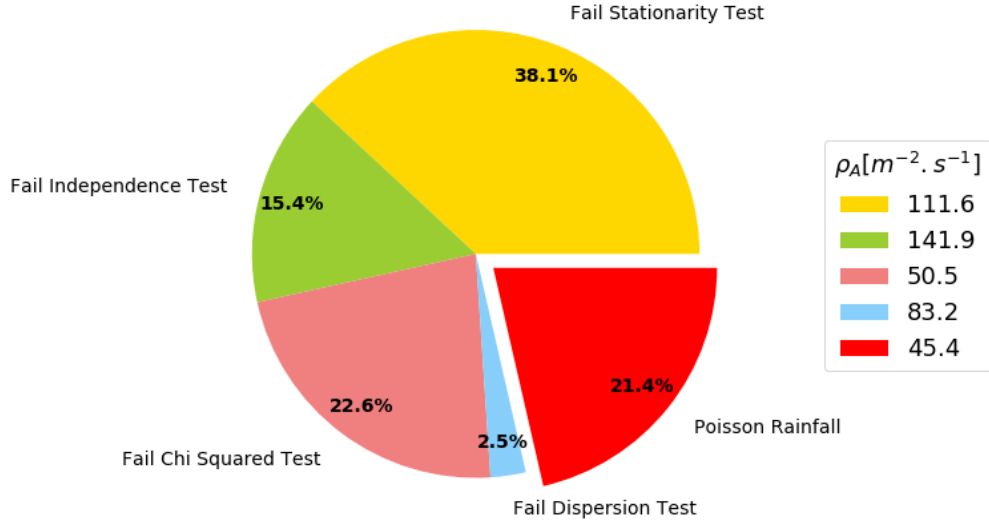


Figure 8. The percentage of all rainfall patches, measured by the intervalometer, that fail each of the hierarchical tests as well as the mean rainfall arrival rate for each group. The data is taken from all of the intervalometer sites.

305 clearly show that the mean rainfall arrival rate is a good predictor of 'Poisson-ness'. Patches of rain with high rainfall arrival rates are typically not classified as Poisson, whereas patches of rainfall with low arrival rates are. This can be clearly seen in the top two right hand panels where the rainfall peak does not pass the Poisson tests but the consistent light rainfall, characterised by low rainfall arrival rate, does.

The reason for the high percentage of Poisson rain at the Chole Mjini site is that the rainfall over this period is dominated by consistent rainfall with a low arrival rate. The signature of this storm can also be seen in the other sites that were online during this period. For example, in the beginning part of the record for Pole Pole, top left (zoomed in top right). And also in both bottom panels between the 01/05/2018 and the 04/05/2018. This consistent light strati-form type rainfall is quite atypical for the rainfall record as a whole. The time series for Pole Pole (top left) and Meremeta (bottom right) show that the record is dominated by intermittent sharp peaks of mostly non Poisson rainfall followed by dry spells. This is the typical pattern for convective rainfall and is consistent with the 'lived' experience during the monsoon.

The disdrometer drop size measurements can be used to characterise Poisson and non-Poisson rainfall patches further and are presented in figure 10. The trend in mean drop size with rainfall arrival rate for Poisson and non Poisson rain is presented in the top panel. This shows that Poisson Rain is characterised by low arrival rates. No Poisson rain is found at $\rho_A > 1500 [m^{-2} \cdot s^{-1}]$. The data also shows a positive correlation between the mean drop sizes and the arrival rate.

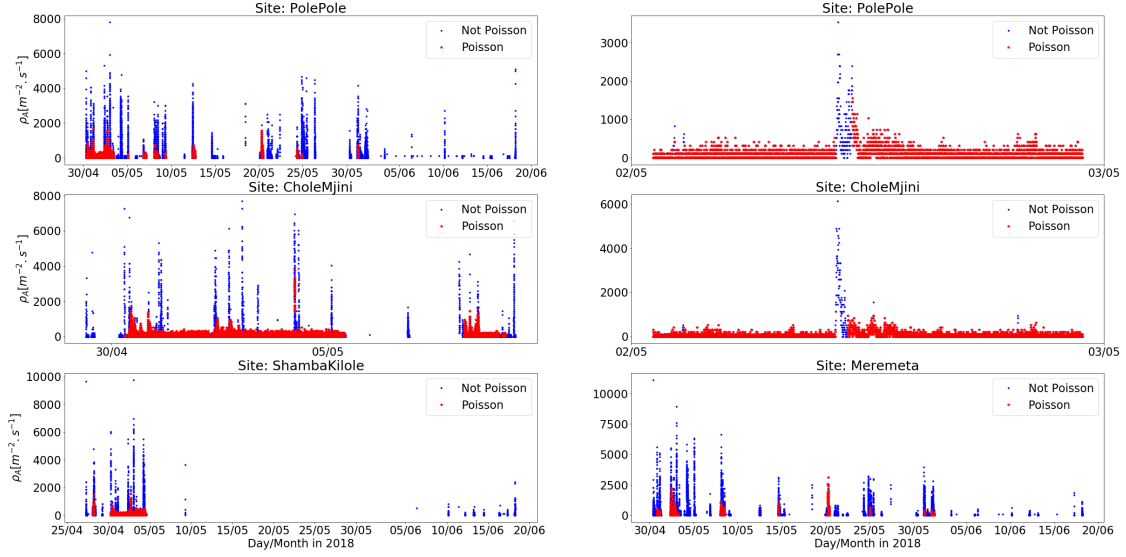


Figure 9. The rainfall record of Pole Pole and Chole Mjini showing an 'atypical' storm characterised by consistent light rainfall rates with little fluctuation over an extended period is presented in the top two rows. This can be clearly seen in the two middle panels of the figure. The time series of rainfall arrival rates clearly show that the mean rainfall arrival rate is a good predictor of 'Poisson-ness'. Patches of rain with high rainfall arrival rates are typically not classified as Poisson, whereas patches of rainfall with low arrival rates are. This can be clearly seen in the top two right hand panels where the rainfall peak does not pass the Poisson tests but the consistent light rainfall, characterised by low rainfall arrival rate, does. This consistent light strati-form type rainfall is quite atypical for the rainfall record as a whole. The time series for Pole Pole (top left) and Meremeta (bottom right) show that the record is dominated by intermittent sharp peaks of mostly non Poisson rainfall followed by dry spells.

320 In the middle panel of figure 10 for each data point the reason for failing to be classified as Poisson rain is also presented. This panel also clearly shows that Poisson rain is found almost entirely at the bottom of the arrival rate range, $\rho_A \leq 600[m^{-2}.s^{-1}]$. As was seen for the intervalometer. This range of rainfall arrival rates contributes little to the total rainfall, 69% of all drops fall in this range but only contribute 16% to total rainfall. Data greater than $2100[m^{-2}.s^{-1}]$ exclusively fail the stationarity and independence tests. This rainfall is therefore characterised by correlations between drop counts and fluctuations in the mean arrival rate at scales smaller than 2-22 minutes. At arrival rates between 700 and $1300[m^{-2}.s^{-1}]$ the rainfall is a mixture of Poisson rain and mostly patches of rainfall that fail the χ^2 . Data that fail the χ^2 test are patches of stationary rainfall with uncorrelated fluctuations about the mean. However the data are over or under dispersed compared to the expected Poisson value of 1 and therefore do not match the Poisson distribution. Mostly, this data is over-dispersed, i.e. the variance is greater than expected by Poisson statistics. As arrival rate increases to between 1400 and $2000[m^{-2}.s^{-1}]$, a higher proportion of rainfall

325

330 (in the sub-range) fails the stationarity and independence tests indicating that rainfall is becoming more and more dynamic (rapid changes in the mean and correlations between drop counts).

In the bottom panel trends in the mean drop size for Poisson and non-Poisson rain are presented. The expected mean drop size of the parameterisation at each arrival rate is also shown. The expected drop sizes are a slight over-estimate of the observed drop sizes, although they are well within the standard error. The overall agreement between the expected and observed drop
335 sizes is quite good and in particular, over the region between 500 and 2500 $m^{-2}.s^{-1}$, which contributes most to the total rainfall amount. This region accounts for 63% of the total rainfall amount. The parameterisation overestimates most of the drop sizes at arrival rates greater than 2500 $m^{-2}.s^{-1}$, however the data becomes quite sparse at higher arrival rates. The positive trend in mean drop size expected by the parameterisation is not as clear for the Poisson data as the non-Poisson data. At arrival rates less than 700 $m^{-2}.s^{-1}$ the Poisson mean drop sizes are larger than the parameterisation and non-Poisson values and at arrival
340 rates greater than 700 $m^{-2}.s^{-1}$ the opposite is the case.

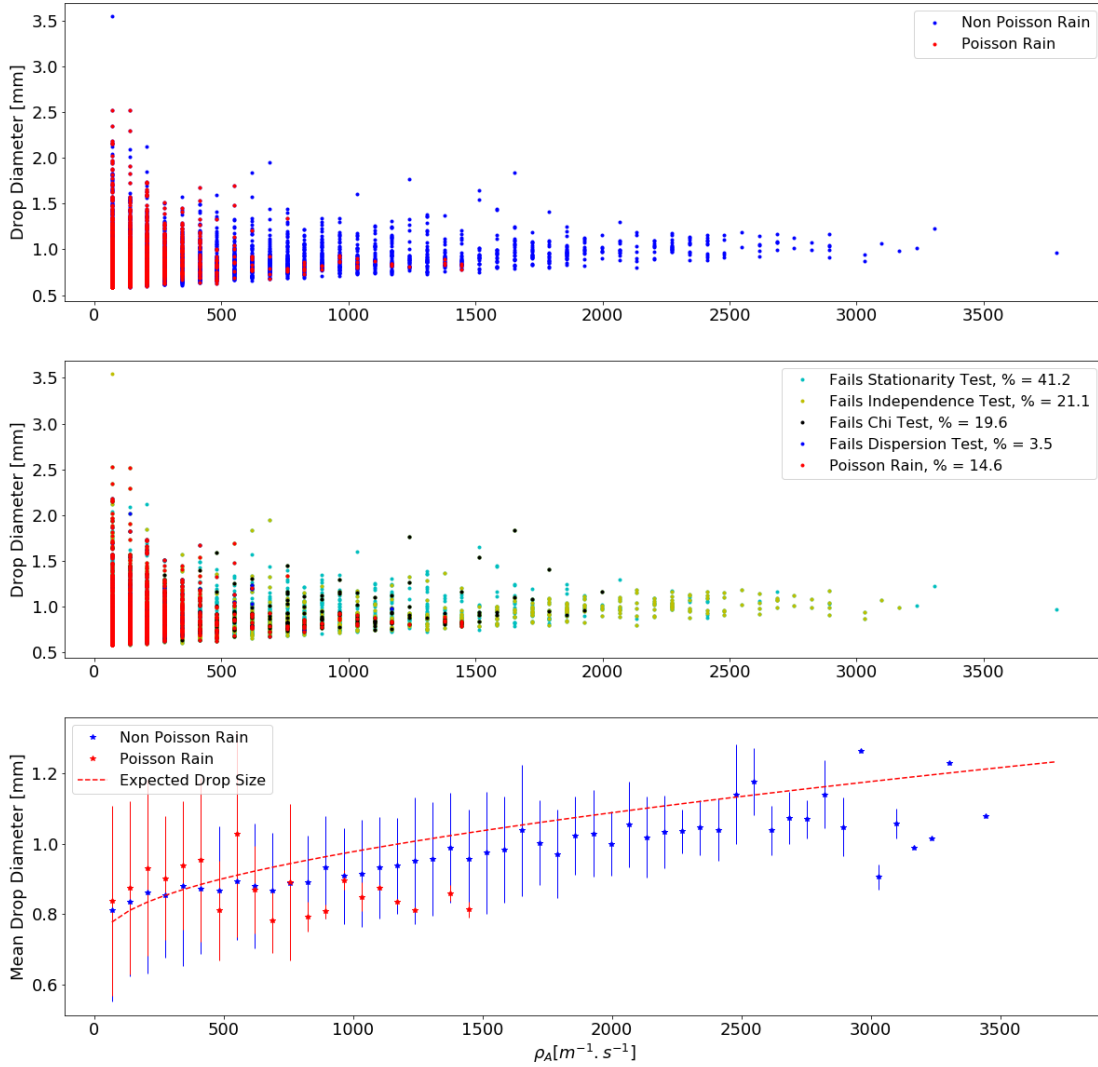


Figure 10. Trends in mean drop size for Poisson and non-Poisson rain are presented as well as the percentage of drops that fail to pass each of the tests for "Poisson-ness". The top panel differentiates between Poisson and non-Poisson rain. The middle panel is further subdivided to show which of the Poisson tests each data point fails. The bottom panel shows the observed mean drop sizes with standard error bars for Poisson (in red) and non-Poisson drops (in blue) as well as the parameterised values (red-dashed line).

5 Discussion

5.1 Rainfall Rates

Accurate estimates of total rainfall can be derived using Marshall and Palmer (1948)'s parameterisation with no adjustment from disdrometer arrival rate measurements. This is because the expected mean drop size of the parameterisation shows good agreement with the observed mean drop sizes. I.e. it is within the std error of the observed mean. In particular, the expected and observed values match quite closely over the range of rainfall arrival rates that contribute most to the total rainfall (63% of total rainfall occurs between 500-2500 $m^{-2}.s^{-1}$). The parameterisation under-estimates the observed mean drop size at low arrival rates $\rho_A \leq 500$ in comparison to observed values. However, it is known that impact disdrometers underestimate the number of small drops and the number of drops in general due to the truncation of drops below the detection limit. Therefore, this difference between the parameterisation and the observed values could be a result of under-reporting of small drops by the instrument. This leads to underestimates of rainfall at low arrival rates. The parameterisation also overestimates the mean size of drops at high arrival rates which leads to over-estimates of the rainfall at high arrival rates. The key point is that Marshall and Palmer (1948)'s parameterisation provides a good estimate of observed mean drop sizes and consequently accurate rainfall estimates can be derived.

This is not the case for the intervalometer estimates of rainfall. The intervalometer results in large over-estimates (by a factor of approximately 3) of the total rainfall amount. This is because the intervalometer registers higher arrival rates during each rainfall event at Pole Pole in comparison to the disdrometer. The intervalometer has a smaller sensor area and a larger D_{min} value than the disdrometer. It should not register higher arrival rates. The possible reasons for the overestimation are, splashing from the intervalometer housing onto the sensor during intense rainfall events, spurious drops due to an electromagnetic signal or physical interference with the sensor, the minimum detectable drop diameter is actually smaller than 0.8mm. Comparison of the rainfall arrival rate records for the disdrometer and intervalometer, for example in figure 5, show that when the intervalometer senses rain, so too does the disdrometer and vice versa. Spurious drops from an interfering signal would also be expected to register outside the rainfall periods. This is not observed. Throughout the observational period and during all rainfall events the intervalometer registers a higher rainfall arrival rate than the disdrometer. I.e. the intervalometer overestimates are not constrained to intense rainfall periods. These two findings indicate that spurious drops and splashing are unlikely causes for the higher arrival rates registered by the intervalometer. It is most likely that the parameter D_{min} was poorly determined and the intervalometer registers drops that are smaller than 0.8mm. The overestimation of rainfall occurs because the parameterisation expects a much larger mean drop size than what is likely observed by the intervalometer.

Since the intervalometer and the disdrometer employ a similar sensor it is reasonable to assume that the drop sizes observed by the intervalometer are of a similar size to those observed by the disdrometer. Using this assumption the intervalometer results are calibrated by the expected values of the mean drop size for the disdrometer. This results in accurate rainfall rates for the intervalometer compared to the tipping bucket (within 5 %) for the entire experiment. This indicates that the actual minimum detectable drop size for the intervalometer is most likely closer to 0.6mm than 0.8mm.

Both forms of the calibration also result in good estimates at another intervalometer site approximately 1 km away (within 9
 375 % of the tipping bucket value). This indicates that the observed mean drop size and therefore the DSD is reasonably stable over
 scales of 1 km. The calibration also gives good results outside the period of time when the disdrometer was online. The last
 rainfall estimate from the intervalometer is approximately 1 month later than the last measurement by the disdrometer. This
 indicates that the DSD is also relatively stable over the entire two month period of the experiment. The derivation of reasonably
 accurate rainfall measurements with both the disdrometer and the intervalometer indicates that Marshall and Palmer's (1948)
 380 parameterisation of the DSD is a good approximation of the observed DSD over the period of the experiment. The interval-
 ometer also shows good potential for being used to derive estimates of the parameter β, D_{min} of the drop size distribution.
 Using only 22 data points it was possible to estimate $\beta = 0.37, D_{min} = 0.53$. More work, with a larger data-set is necessary to
 fully assess the validity of using intervalometer measurements for deriving estimates of the DSD parameters but this first step
 shows good promise.

385 5.2 Testing the Poisson Hypothesis

The results show that the majority of rainfall does not comply with the Poisson Homogeneity assumption. Over all the sites only
 22.5 % of all the raindrops observed by the intervalometers can be reasonably assumed to behave according to Poisson statistics.
 For the disdrometer only 15% of the rainfall behaves according to Poisson statistics. The majority of this Poisson rainfall is to
 be found in a series of "atypical" rainfall events. These events are atypical because they are characterised by consistent periods
 390 of light rainfall that have a duration of up to several hours interspersed with sharper peaks of higher intensity. The rest of the
 rainfall record is characterised by short intense showers with high arrival rates preceded and followed by dry periods. It seems
 that rainfall can be divided into two types over the experimental period. Consistent light rain which is most often classified as
 Poisson and short, intense showers that are never classified as Poisson.

The results of the tests indicate that high arrival rates are indicative of rainfall which has a fluctuating mean on very short time
 395 scales (< 2 min in some cases). Rainfall with high arrival rates is also characterised by correlations between drop counts on
 very short time scales. Almost all of the rainfall that contributes most to the total rainfall amount does not exhibit characteristics
 that are consistent with Poisson statistics. One would then expect that estimates of rainfall based on a parameterisation that has
 been defined independently of the size of a reference volume, thus implying an assumption of homogeneity, would not return
 good results. This is not the case.

400 Estimates of rainfall are good and more surprisingly the trend in mean drop size with increasing rainfall arrival rate is not
 only consistent with expected values derived from the parameterisation but also appears to be mostly captured by *non-Poisson*
 rainfall. The trend in the mean drop size of Poisson rainfall with increasing arrival rate is much less clear. This would imply
 that estimates of rainfall derived from an exponential parameterisation of the DSD would be less accurate over the patches
 of rainfall that contain Poisson rain as opposed to patches of rainfall that contain non-Poisson rain. In figure 11 two different
 405 rainfall patches of a similar total rainfall amount but very different arrival rate profiles are compared. One event contains a
 significant proportion of Poisson rain and the other contains no Poisson rain. The figure clearly shows that the quality of the
 rainfall estimate is much worse for the rainfall event that contains Poisson rain. In that event rainfall is under-estimated by

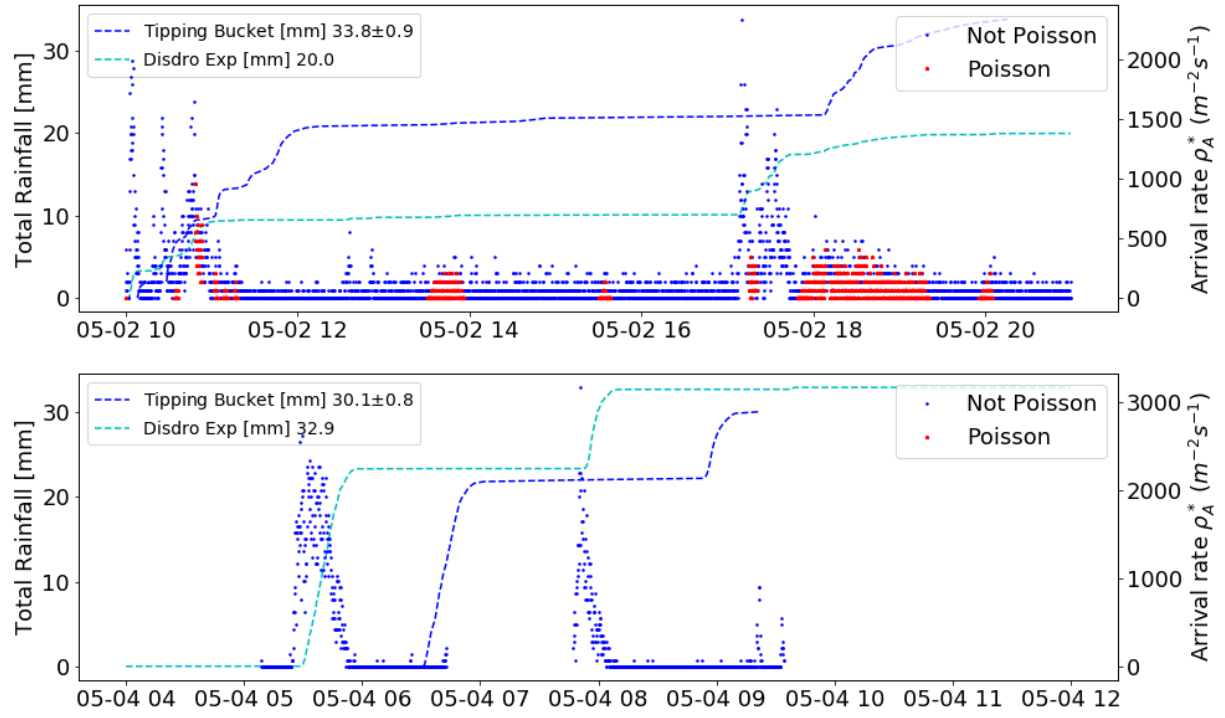


Figure 11. The performance of the rainfall parameterisation over a period of rainfall with a high proportion of "Poisson Rain" (top panel) compared to a period of rainfall with a similar total rainfall amount but with no "Poisson Rain" (bottom panel). The disdrometer estimates are plotted against the Tipping bucket values and the rainfall arrival rate in both panels.

approximately 41%. In the rainfall event with no Poisson rain, the parameterised estimate is within 10% of the tipping bucket value. This seems to indicate that the presence of Poisson rainfall leads to worse rainfall estimates. However, the rainfall event with Poisson rain also contains a significantly higher proportion of light rainfall in general (both Poisson and not) compared to the rainfall event with no Poisson rain. It is known that impact disdrometers underestimate the numbers of small drops and therefore the rainfall rate at low rainfall arrival rates. So, whilst the figure does show that rainfall estimates are worse when there is Poisson rainfall this cannot be de-tangled from the fact that rainfall estimates in general are also worse when arrival rates are low. More work needs to be done in order to understand if the poor rainfall estimates are due to Poisson rain or are simply an artefact of the measuring instrument. However, this research does show that the compliance with the Poisson homogeneity hypothesis is not necessary for deriving accurate rainfall estimates.

One of the criticisms that arises with the statistical tests employed in this research is that the tests are less strict with smaller sample sizes and also at lower arrival rates. This could bias the results such that rainfall with low arrival rates is more likely to pass all of the tests. This was understood by Cornford (1967) and led to his simple sampling criterion requiring 23 drops per bin size. This criterion is not fulfilled in this study. However as is pointed out by Jameson and Kostinski (1997); Jameson and Kostinski (1998), rainfall conditions are changing rapidly, sometimes at scales less than 2 minutes. The presence of these fine structures within rainfall would be obscured by larger sampling windows. Furthermore sampling across such structures with different means may actually lead to increased uncertainty in the mean. This increased uncertainty in the mean over an entire rainfall event would make it impossible to test the Homogeneous Poisson assumption because rainfall is very rarely stationary over longer time periods. Therefore in such cases the sampling criteria need to be adjusted to account for the patch size. In this research it was decided to treat τ as a moving window to increase the effective number of samples and account for the small sample size. In this way the same tests are run on each of the drop counts many times, providing more robust and reliable results.

6 Conclusions

This research leads to the following conclusions.

1. The majority of rainfall that was observed is not consistent with Poisson statistics on observation scales from 2-22 minutes. The observed Poisson rainfall is characterised by low mean rainfall arrival rates. No Poisson rain is observed with $\rho_A > 1500m^{-2}.s^{-1}$.
2. The majority of the Poisson rainfall can be associated with a series of storms over a three day period that are atypical in comparison to the entire observed rainfall period. These storms are characterised by long periods of light stratiform type rainfall, most likely caused by a large scale synoptic forcing. The rest of the rainfall record is mostly comprised of convective showers.
3. The homogeneous Poisson assumption does not apply for the majority of rainfall observed in this study. Rainfall shows correlations between drop counts and changes in the mean at scales as small as 2 min. It is possible that rainfall is homogeneously distributed at smaller time scales but these would be so small as to invalidate the very concept of a drop size distribution.
4. Despite the apparent invalidity of the Homogeneous Poisson assumption, plots of mean drop sizes against rainfall arrival rate reveal that the expected mean drop sizes from Marshall and Palmer (1948) parameterisation shows good agreement with observed values both over spatial scales of 1km and a temporal period of 2 months.

- 450 5. Total cumulative rainfall estimates derived from the disdrometer drop counts are within the standard error of the total rainfall amount measured by a co-located tipping bucket over the same time period.
- 455 6. The intervalometers at both tipping bucket sites give large over estimates of the total rainfall. This is most likely due to a poor calibration of the parameter D_{min} . The actual D_{min} is most likely close to 0.6mm. Constraining the intervalometer arrival rates by the observed mean drop sizes results in rainfall estimates that are within within 5-10% of tipping bucket measurements. The form of the constraint relationship is the parameterisation used for the disdrometer measurements. The accuracy of rainfall estimates is determined by the accuracy of the DSD parameterisation.
- 460 7. It is possible to determine reasonable rainfall estimates using an intervalometer. It is also likely that the intervalometer can be used in conjunction with co-located rain gauges to give good estimates of mean drop sizes and therefore the parameters of the exponential DSD. In turn this may improve satellite-derived rainfall estimates. Due to its low cost, the instrument shows good potential for being deployed in Africa to alleviate the observational crisis.

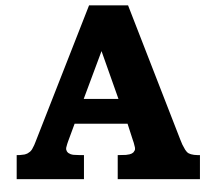
Competing interests. No competing interests are present

465 *Acknowledgements.* The work leading to these results has received funding from the European Community's Horizon 2020 Programme (2014-2020) under grant agreement No. 776691 (TWIGA). The opinions expressed in the document are of the authors only and no way reflect the European Commission's opinions. The European Union is not liable for any use that may be made of the information. Jan Jaap Pape is acknowledged for his preliminary work in building, designing and testing the early generations of the intervalometer. Thanks go to Disdro in Delft for use of their workshop and help in building the newer version of the intervalometer. Special thanks to the hotels located on Mafia Island (Didimiza Guest House, Meremeta Lodge, Eco Shamba Kilole Lodge, Kinasi Lodge, Pole Pole Bungalows and the Mafia 470 Island Lodge) and Chole Island (Chole Mjini Treehouse Lodge) for allowing access to their land and support in setting up the experiment.

References

- Guide to meteorological instruments and methods of observation, Tech. rep., World Meteorological Organisation, 2014.
- Robust Climate Observation in Africa and Low Resource Countries: Developing Practicable Public Private Partnership Models, Trans-African Hydro-Meteorological Observatory (TAHMO), 2017.
- 475 Abdrabo, M., Essel, A., Lennard, C., Padgham, J., and Urquhart, P.: Africa. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., Cambridge, United Kingdom and New York, NY, USA, 2014.
- Atlas, D. and Ulbrich, C. W.: Path- and Area-Integrated Rainfall Measurement by Microwave Attenuation in the 1–3 cm Band, *Journal of Applied Meteorology*, 16, 1322–1331, [https://doi.org/10.1175/1520-0450\(1977\)016<1322:paairm>2.0.co;2](https://doi.org/10.1175/1520-0450(1977)016<1322:paairm>2.0.co;2), 1977.
- 480 Battan, L. J.: Radar observation of the atmosphere, University of Chicago Press, 1973.
- Boko, M., Niang, I., and Nyong, A.: Africa. In: Climate change adaptation and vulnerability: contribution of working group II to the IV assessment report of the IPCC panel on climate change., Tech. rep., Cambridge, United Kingdom and New York, NY, USA, 2007.
- Bringi, V. N., Chandrasekhar, V., Hubbert, J., Gorgucci, E., Randeu, W. L., and Schoenhuber, M.: Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis, *Journal of Atmospheric Science*, 60, 354–365, 2003.
- 485 Ciach, G. J.: Local Random Errors in Tipping-Bucket Rain Gauge Measurements, *Journal of Atmospheric and Ocean Technology*, [https://doi.org/10.1175/1520-0426\(2003\)20<752:LREITB>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2), 2003.
- Cornford, G. S.: Sampling errors in measurements of rain- drop and cloud droplet concentration, *Meteorology Magazine*, 96, 271–282, 1967.
- Feingold, G. and Levin, Z.: The Lognormal Fit to Raindrop Spectra from Frontal Convective Clouds in Israel, *Journal of Climate and Applied Meteorology*, 25, 1346–1363, [https://doi.org/10.1175/1520-0450\(1986\)025<1346:tlftrs>2.0.co;2](https://doi.org/10.1175/1520-0450(1986)025<1346:tlftrs>2.0.co;2), 1986.
- 490 Feller, W.: An introduction to probability theory and its applications, Wiley, 2010.
- Greatrex, H., Hansen, J., Garvin, S., Diro, R., Blakeley, S., Le Guen, M., Rao, K., and Osgood, D.: Scaling up index insurance for smallholder farmers: Recent evidence and insights., Tech. Rep. 14, CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark, 2015.
- Gunn, R. and Kinzer, G. D.: The Terminal Velocity Of Fall For Water Droplets In Stagnant Air, *Journal of Meteorology*, 6, 243–248, [https://doi.org/10.1175/1520-0469\(1949\)006<0243:ttvoff>2.0.co;2](https://doi.org/10.1175/1520-0469(1949)006<0243:ttvoff>2.0.co;2), 1949.
- 495 Guyot, A., Pudashine, J., Protat, A., Uijlenhoet, R., Pauwels, V. R. N., Seed, A., and Walker, J. P.: Effect of disdrometer type on rain drop size distributioncharacterisation: a new dataset for Southeastern Australia, *Hydrology and Earth System Sciences Discussions*, p. 1–39, <https://doi.org/10.5194/hess-2019-277>, 2019.
- Hershey, J. R. and Olsen, P. A.: Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models, *Institute of Electrical and Electronic Engineers*, <https://doi.org/10.1109/ICASSP.2007.366913>, 2007.
- 500 Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., Le, M., Chandrasekar, V., and Kubota, T.: GPM/DPR Level-2 Algorithm Theoretical Basis Document, Tech. rep., 2017.
- Jameson, A. R. and Kostinski, A. B.: Fluctuation Properties of Precipitation. Part II: Reconsideration of the Meaning and Measurement of Raindrop Size Distributions, *Journal of the Atmospheric Sciences*, 55, 283–294, [https://doi.org/10.1175/1520-0469\(1998\)055<0283:fpoppi>2.0.co;2](https://doi.org/10.1175/1520-0469(1998)055<0283:fpoppi>2.0.co;2), 1998.
- 505 Johnson, R. W., Kliche, D. V., and Smith, P. L.: Comparison of Estimators for Parameters of Gamma Distributions with Left-Truncated Samples, *Journal of Applied Meteorology and Climatology*, 50, 296–310, <https://doi.org/10.1175/2010jamec2478.1>, 2011.

- Joss, J. and Waldvogel, A.: Raindrop Size Distribution and Sampling Size Errors, *Journal of the Atmospheric Sciences*, 26, 566–569, [https://doi.org/10.1175/1520-0469\(1969\)026<0566:rsdass>2.0.co;2](https://doi.org/10.1175/1520-0469(1969)026<0566:rsdass>2.0.co;2), 1969.
- 510 Kostinski, A. B. and Jameson, A. R.: Fluctuation Properties of Precipitation. Part I: On Deviations of Single-Size Drop Counts from the Poisson Distribution, *Journal of the Atmospheric Sciences*, 54, 2174–2186, [https://doi.org/10.1175/1520-0469\(1997\)054<2174:fpoppi>2.0.co;2](https://doi.org/10.1175/1520-0469(1997)054<2174:fpoppi>2.0.co;2), 1997.
- Lavergnat, J. and Golé, P.: A Stochastic Raindrop Time Distribution Model, *Journal of Applied Meteorology*, 37, 805–818, [https://doi.org/10.1175/1520-0450\(1998\)037<0805:asrtdm>2.0.co;2](https://doi.org/10.1175/1520-0450(1998)037<0805:asrtdm>2.0.co;2), 1998.
- 515 Lovejoy, S. and Schertzer, D.: Fractals, Raindrops and Resolution Dependence of Rain Measurements, *Journal of Applied Meteorology*, 29, 1167–1170, [https://doi.org/10.1175/1520-0450\(1990\)029<1167:frardo>2.0.co;2](https://doi.org/10.1175/1520-0450(1990)029<1167:frardo>2.0.co;2), 1990.
- Marshall, J. S. and Palmer, W. M.: Uncertainties of GPM DPR Rain Estimates Caused by DSD Parameterizations, *Journal of Meteorology*, 5, 165–166, 1948.
- Munchak, S. J. and Tokay, A.: Retrieval of Raindrop Size Distribution from Simulated Dual-Frequency Radar Measurements, *Journal of*
520 *Applied Meteorology and Climatology*, 47, 223–239, <https://doi.org/10.1175/2007jamc1524.1>, 2008.
- Otto, F. E. L., Boyd, E., Jones, R. G., Cornforth, R. J., James, R., Parker, H. R., and Allen, M. R.: Attribution of extreme weather events in Africa: a preliminary exploration of the science and policy implications, *Climatic Change*, 132, 531–543, <https://doi.org/10.1007/s10584-015-1432-0>, 2015.
- Pape, J. J.: Drop Counter: Testing the performance of a simple acoustic disdrometer to measure rainfall, 2018.
- 525 Sevruk, B.: Correction of Precipitation Measurements:, Tech. rep., Swiss Federal Institute of Technology, 1985.
- Smith, J. A.: Marked Point Process Models of Raindrop-Size Distributions, *Journal of Applied Meteorology*, 32, 284–296, [https://doi.org/10.1175/1520-0450\(1993\)032<0284:mppmor>2.0.co;2](https://doi.org/10.1175/1520-0450(1993)032<0284:mppmor>2.0.co;2), 1993.
- Tokay, A. and Short, D. A.: Evidence from Tropical Raindrop Spectra of the Origin of Rain from Stratiform versus Convective Clouds, *Journal of Applied Meteorology*, 35, 355–371, [https://doi.org/10.1175/1520-0450\(1996\)035<0355:efterso>2.0.co;2](https://doi.org/10.1175/1520-0450(1996)035<0355:efterso>2.0.co;2), 1996.
- 530 Tokay, A., Kruger, A., and Krajewski, W. F.: Comparison of Drop Size Distribution Measurements by Impact and Optical Disdrometers, *Journal of Applied Meteorology*, 40, 2083–2097, [https://doi.org/10.1175/1520-0450\(2001\)040<2083:codsdm>2.0.co;2](https://doi.org/10.1175/1520-0450(2001)040<2083:codsdm>2.0.co;2), 2001.
- Uijlenhoet, R. and Stricker, J. N. M.: A consistent rainfall parameterization based on the exponential drop size distribution, *Journal of Hydrology*, 218, 101–127, 1999.
- Uijlenhoet, R., Stricker, J., Torfs, P., and Creutin, J.-D.: Towards a stochastic model of rainfall for radar hydrology: testing the
535 poisson homogeneity hypothesis, *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 24, 747–755, [https://doi.org/10.1016/s1464-1909\(99\)00076-3](https://doi.org/10.1016/s1464-1909(99)00076-3), 1999.
- Ulbrich, C. W.: Numerical Variations in the Analytical Form of the Raindrop Size Distribution, *Journal of Climate and Applied Meteorology*, 22, 1764–1775, 1983.
- Washington, R., Harrison, M., Conway, D., Black, E., Challinor, A., Grimes, D., Jones, R., Morse, A., Kay, G., and Todd, M.: African Climate
540 Change: Taking the Shorter Route, *Bulletin of the American Meteorological Society*, <https://doi.org/10.1175/BAMS-87-10-1355>, 2006.



README for the python code files

This appendix is a README for the python code files that were developed during the course of the MSc research. Note, the actual python code is not contained within this document. It has been submitted as supplementary material, along with the raw and processed data, to the TU Delft education repository and can be found at <https://repository.tudelft.nl>, by searching for the title of this MSc thesis.

A.1. Overview of the code

In this README each of the various python scripts that were developed during the course of this MSc research are explained. The order in which they should be used is also laid out. The research resulted in X different python scripts being developed. Each script completes a different set of tasks. Each of the scripts could conceivably be combined into one long script but this would be clumsy to run and debug.

Note, it is recommended to download all the raw data (tipping bucket, intervalometer and disdrometer data) from <https://repository.tudelft.nl/>. It is recommended to keep the same folder organization for the raw data files as in the zipped file. I.e. do not change folder names or move folders within the downloaded data. Save the downloaded data to a chosen folder and specify that folder's path.

The scripts, listed in the order they are used to analyse the data, are as follows:

1. import_DC.py
2. import_disdro.py
3. import_tb.py
4. continuous_record_dc.py
5. continuous_record_disdro.py
6. stations_online.py
7. Poisson_Testing.py
8. PoissonAnalysis.py
9. RainfallAnalysis.py

And the associated functions, also listed in the order that they are used:

1. poisson_test.py
2. poisontest.py
3. seperate_rainfall_events.py

4. mean_drop_sizes.py

A more detailed description of each script and function is presented in the next section.

A.2. Description of each script and function

A.2.1. Script Name: import_DC.py

Overview

This script reads all of the raw intervalometer txt files from each of the sites and processes the data. The raw data is in 4 forms all mixed into one txt file; a millisecond Unix timestamp at the start of each txt file, timestamps (in millisecond Unix time) for each time a drop is registered by the sensor, a check timestamp every 10 minutes so that you know that the intervalometer is online even when it is not raining and a voltage stamp (depending on the version of the Arduino software installed on the intervalometer). Version 6 includes voltage readings and Version 5 does not.

The script sorts the drop data from the check and voltage data and records the start and end times of each txt file. The end time is taken as the time of the last drop. Unix time is converted to date-time in UTC. The script also imports a manual record of when the intervalometer was being physically handled. E.g. working around the sensor to download data etc. The script deletes all drops registered within the manual record windows since these are spurious drops from touching the sensor. All the start and end times of each of the txt files are used to determine a continuous record for the instrument. I.e. the time period when the instrument was online and registering rainfall. Finally the continuous record data, drop arrival data and check/voltage data are each saved to their own txt file.

Usage

Modify the root_path variable within this script so that it points to the folder where all the data has been saved. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script -s 'Name_of_site'`

For example: `python '/Users/didierdevilliers/Documents/TU_Delft/Graduation/Python_Scripts/import_DC.py' -s 'Didimiza'`

Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

The excel file containing the manual record of working around the sensors and all the raw intervalometer data txt files.

Outputs

The script produces three txt files which contain the voltage data, processed drop arrival data and check/voltage data. These files are saved to sub-folders within the root_path folder.

A.2.2. Script Name: import_disdro.py

Overview

This script reads the raw disdrometer csv file and removes any spurious drops from within the manual record windows that may have been caused by touching the sensor. The 'cleaned' data is saved to a txt file.

Usage

Modify the `root_path` variable within this script so that it points to the folder where all the data has been saved. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

The excel file containing the manual record of working around the sensors and the csv file containing the raw disdrometer data.

Outputs

One txt file containing the cleaned drop data from the disdrometer.

A.2.3. Script Name: `import_tb.py`

Overview

This script imports the raw tipping bucket data in csv files from Shamba Kilole, MIL1 and Pole Pole and removes any spurious tips from within the manual record windows. The results of the field calibration are also applied to convert from tips to mm of rainfall. The cleaned and processed rainfall amounts as well as the tips are saved to a txt file.

Usage

Modify the `root_path` variable within this script so that it points to the folder where all the data has been saved. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

The csv files of raw tipping bucket data from the three sites and the excel file containing the manual record of working around the sensors.

Outputs

A txt file containing the rainfall tips and volume of each tip (determined by a field calibration experiment).

A.2.4. Script Name: `continuous_record_dc.py`

Overview

This script reads all the processed intervalometer drop data txt files that were generated with the `import_DC.py` script as well as the files containing the continuous record (periods of online operation) for each intervalometer. All the drop data within the continuous record periods is merged into one data frame for each site along with an indication of which continuous record period a drop corresponds to. The complete continuous record of drops for each site is then re-sampled into 10 second time bins. The re-sampled drop data as well as the continuous record of drops are saved to separate txt files for each site. I.e. this script takes many different txt files of drop data from each site and combines them into two txt files, one with data that has been re-sampled to 10s time bins and one with the original timestamps. The script also makes some plots with the drop data for each site.

Usage

Modify the `root_path`, `drop_path` and `cr_path` variables within this script so that they point to the relevant folders. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

Processed drop data for each intervalometer site as well as the continuous record data for each intervalometer site.

Outputs

Two txt files for each intervalometer site, one containing the entire record of continuous drop data for the site and the other containing the same record but re-sampled into 10s time bins.

A.2.5. Script Name: `continuous_record_disdro.py`

Overview

This script reads the processed disdrometer data txt file that was generated with the `import_disdro.py` script and deletes data after the malfunction date. The remaining data is re-sampled into 10s time bins. For each bin some basic statistical indices of the drop sizes (mean, median, var etc) are calculated. The processed data are saved to a txt file. The script also makes some plots of the drop arrival time series.

Usage

Modify the `root_path` and `disdro_path` within this script so that they point to the relevant folders. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

Processed drop data from the disdrometer.

Outputs

One txt file containing re-sampled drop data and basic statistical indices for each 10s bin.

A.2.6. Script Name: `stations_online.py`

Overview

This script reads the processed tipping bucket data from Pole Pole generated by `import_tb.py`, the processed disdrometer data generated by `continuous_record_disdro.py` and the continuous record data for each intervalometer site generated by `import_DC.py` and uses this to generate a plot showing the periods within the data record when the different instruments are online or offline in comparison to one another. The plot is saved to a specified path.

Usage

Modify the `root_path` variable and the path where the figure is saved so that they point to the relevant folders. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

Continuous record data, processed tipping bucket data and processed disdrometer data.

Outputs

A plot saved in a png file.

A.2.7. Script Name: Poisson_Testing.py

Overview

This script reads the re-sampled disdrometer and intervalometer data (for each site) and then performs some tasks. It first separates the raindrop record into distinct rain events, using the function `seperate_rain_events.py`, determined by a dry period of greater than 1 hour between consecutive drops. The auto correlation of each rainfall event at increasing lag times is calculated. The lag time at which the auto-correlation drops below $\frac{1}{e}$ is defined as τ . A check is performed to determine if $t \ll \tau \ll T$, where $t = 10s$ and T is the length of the rainfall event. If the rainfall event passes this 'Kostinski' criterion it is labelled a 'Kostinski' storm. The script then passes all the Kostinski storms to another function called `poisson_test.py`. This function performs all of the hierarchical tests for 'Poisson-ness' on each of the Kostinski storms for each of the sites. The distinct rain events, Kostinski storms and the results of the Poisson tests are all saved to their own txt file. This analysis is performed for the intervalometer data at each site as well as the disdrometer data.

Usage

Modify the `root_path` and `disdro_path` within this script so that they point to the relevant folders. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

The re-sampled (10s) disdrometer and intervalometer data and two functions (`seperate_rain_events.py` and `poisson_test.py`).

Outputs

Three txt files for each intervalometer site as well as the disdrometer site. The txt files contain, the separate rain events, the Kostinski storms and the results of the Poisson tests.

A.2.8. Script Name: PoissonAnalysis.py

Overview

This script reads in many data files; the results of the Poisson tests, the Kostinski storms, the separate rain events, the re-sampled drop data and the un-resampled drop data for the disdrometer and all the intervalometer sites. Several different analyses are performed on the data.

- All the single drops within the un-resampled record are classified according to the results of the Poisson tests
- Basic statistical indices are calculated for these data groups and plots are made
- Expected mean drop sizes are calculated using the function: `exp_drop_size.py`

- These are compared with observed values and more plots are made
- Time series plots are made for Poisson vs non-Poisson rainfall
- Trends in the % of large drops are plotted

All the plots are saved to a prescribed path.

Usage

Modify the `root_path` within this script so that it points to the relevant folders. No other changes are necessary. This script must be opened in a python IDE, such as Spyder, and run within that environment.

Inputs

The function `exp_drop_size.py` and the following data for the intervalometer sites and the disdrometer:

- Results of Poisson tests
- Kostinski storms
- The separated rainfall events
- Re-sampled drop data
- Un-resampled drop data

Outputs

This script produces many different figures based on the analyses.

A.2.9. Script Name: RainfallAnalysis.py

Overview

This script reads in many data files; the results of the Poisson tests, the Kostinski storms, the separate rain events, the re-sampled drop data and the un-resampled drop data for the disdrometer and all the intervalometer sites. It also reads in the tipping bucket rainfall data for each of the three tipping bucket sites. The main function of this script is to calculate rainfall rates using the functions, `expon_rain.py` and `rain_adf_final.py`, from the intervalometer and disdrometer re-sampled (10s bins) drop data. The calculated rainfall rates are then compared with the tipping bucket values over the same time period. Several different plots are produced and saved.

Usage

Modify the `root_path` within this script so that they point to the relevant folders. No other changes are necessary. This script can be run from the terminal by calling: `python path_to_script`. Alternatively, you can open the script in a python IDE, such as Spyder, and run it within that environment.

Inputs

This script requires the following functions; `expon_rain.py`, `exp_drop_size.py`, `rain_adj_final.py` and `exp_poly_constants.py`, and the following data files:

- Results of Poisson tests
- Kostinski storms
- The separated rainfall events

- Re-sampled drop data
- Un-resampled drop data
- Tipping bucket rainfall data

Outputs

This script produces many different figures based on the analyses.

A.2.10. Function Name: `seperate_rain_events.py`

Overview

This function takes three arguments; a data frame of re-sampled drop data, a string containing the name of the site and a string specifying the re-sample period (in this case 10s). The function takes the re-sampled drop data and separates it into distinct rainfall events using a criterion of more than 1 hour between consecutive drops. The value of τ for each storm is determined and the 'Kostinski' criterion is applied. The function returns the separated rain events, the Kostinski storms, the T and τ used to determine the Kostinski storms and a continuous record counter.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

This function takes three arguments; a data frame of re-sampled drop data, a string containing the name of the site and a string specifying the re-sample period (in this case 10s).

Outputs

The function returns the separated rain events, the Kostinski storms, the criterion used to determine the Kostinski storms and a continuous record counter.

A.2.11. Function Name: `poisson_test.py`

Overview

This function takes the Kostinski storms and performs a series of tests of them to determine if the rainfall data can reasonably be assumed to comply with the Poisson Homogeneity hypothesis. The tests are performed on a sub-section of each Kostinski storm with length τ , determined in the `seperate_rain_events.py` function. The series of tests that are performed are explained in the methodology section of the paper. This function returns the results of the tests for each sub-section of each Kostinski storm.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The function takes the following arguments: Kostinski storm data, storm number (identifier), τ value, n = number of sub-sections within each storm, continuous record period, re-sample time, site name and the instrument (disdrometer or intervalometer). All of these inputs are calculated in previous scripts or are outputs of previous functions.

Outputs

This function returns the results of the tests for each sub-section of each Kostinski storm in a data frame.

A.2.12. Function Name: `expon_rain.py`

Overview

This function calculates the rainfall rate from the rainfall arrival rate using Marshall and Palmer's (1948) parameterisation. The rainfall rate is calculated using the polynomial constants from the function `exp_poly_constants.py`. The function returns a data frame containing the rainfall rates.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The function takes three arguments, the rainfall arrival rate data, the instrument (disdrometer or intervalometer) and the value of D_{min} . The function also uses another function called `exp_poly_constants.py`.

Outputs

The rainfall rates.

A.2.13. Function Name: `exp_poly_constants.py`

Overview

This function fits a third degree polynomial to the $\rho_A - R$ relationship given by Marshall and Palmer's (1948) parameterisation. It takes only one argument, D_{min} , the minimum detectable drop size and returns the polynomial constants.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The value of D_{min} .

Outputs

The polynomial constants.

A.2.14. Function Name: `exp_drop_size.py`

Overview

This function calculates the expected mean drop size as a function of rainfall arrival rate using the expectation of a left truncated gamma distribution. It first converts the rainfall arrival rate to rainfall rate using the function `exp_poly_constants.py` and then converts the rainfall rate to Λ in order to calculate the expected mean drop size. The function returns the expected drop size for the truncated distribution as well as for the complete distribution.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The function takes three arguments, the rainfall arrival rate data, the instrument (disdrometer or intervalometer) and the value of D_{min} . The function also uses another function called `exp_poly_constants.py`.

Outputs

The mean expected drop size for the truncated gamma and complete gamma distributions.

A.2.15. Function Name: `expon_rain_adj.py`

Overview

This function constrains the rainfall estimates from Marshall and Palmer's (1948) parameterisation by a priori observations of the mean drop sizes from the disdrometer by using the equation derived in the methodology section of the paper for a complete gamma distribution.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The function takes three arguments, the rainfall arrival rate data, the instrument (disdrometer or intervalometer) and the value of D_{min} . The function also uses two other functions, `exp_poly_constants.py` and `exp_drop_size.py`.

Outputs

The corrected rainfall rates.

A.2.16. Function Name: `rain_adj_final.py`

Overview

This function constrains the rainfall estimates from Marshall and Palmer's (1948) parameterisation by a priori observations of the mean drop sizes from the disdrometer by using the equation derived in the methodology section of the paper for a left truncated gamma distribution.

Usage

The function must be imported into the relevant script (as you would import any python module) and can then be called within the script by its name. Note, for the import to work, the function location in your hard drive must be part of the Python PATH.

Inputs

The function takes three arguments, the rainfall arrival rate data, the instrument (disdrometer or intervalometer) and the value of D_{min} . The function also uses two other functions, `exp_poly_constants.py` and `exp_drop_size.py`.

Outputs

The corrected rainfall rates.