

# Probability Distribution of Philippine Daily Rainfall Data

Vanessa Althea B. Bermudez (vbbermudez1@up.edu.ph)<sup>1</sup>, Ariel Bettina B. Abilgos (ababilgos@up.edu.ph)<sup>1</sup>,  
Diane Carmeliza N. Cuaresma (dncuaresma@up.edu.ph)<sup>1</sup>, and Jomar F. Rabajante (jfrabajante@up.edu.ph)<sup>1,\*</sup>

<sup>1</sup>Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños, Laguna  
4031 Philippines

\*Corresponding author: jfrabajante@up.edu.ph

## 1 Abstract

Philippines as an archipelago and tropical country, which is situated near the Pacific ocean, faces uncertain rainfall intensities. This makes environmental, agricultural and economic systems affected by precipitation difficult to manage. Time series analysis of Philippine rainfall pattern has been previously done, but there is no study investigating its probability distribution. Modeling the Philippine rainfall using probability distributions is essential, especially in managing risks and designing insurance products. Here, daily and cumulative rainfall data (January 1961 - August 2016) from 28 PAGASA weather stations are fitted to probability distributions. Moreover, the fitted distributions are examined for invariance under subsets of the rainfall data set. We observe that the Gamma distribution is a suitable fit for the daily up to the ten-day cumulative rainfall data. Our results can be used in agriculture, especially in forecasting claims in weather index-based insurance.

*Keywords:* Philippines, rainfall, precipitation, Gamma distribution, probability, weather risk

## 2 Introduction

Philippines, as an agricultural country, allocates about 32% of its total land area to crop production [1]. The major crops include rice, corn, sugarcane and banana. In crop production, it is essential to consider climate and weather factors that will greatly determine the potential growth and yield of harvests. An example of these factors is rainfall availability.

Erratic rainfall pattern and extreme events could lead to production losses [2, 3]. Being an archipelagic tropical country near the Pacific ocean, Philippines experiences more of the adverse effects of the fast-changing climate [4]. For instance, scarce or excessive rainfall events (e.g. El Niño and La Niña) occur more often and more intense than usual [5]. These situations affect many sectors negatively or worse, these can lead to destruction of valuable assets. These circumstances could also threaten food security, telecommunication, and coastal communities. Due to these consequences, it is necessary to determine rainfall patterns in different parts of

the country. Since rainfall is highly variable, we can study this using probability distributions [6, 7, 8].

When the correct probability distributions for rainfall data are determined, we can easily simulate or forecast rainfall amounts without losing its accuracy and reliability. The simulation can be used for managing water resources, such as watersheds or in rain-fed farms. It can also be employed in computations for weather index-based insurances for crops [9].

Time series analysis of Philippine rainfall data and its seasonality has been previously done [10, 5, 11, 12]. However, there is still no existing published results about the probability distribution of the rainfall pattern in the Philippines. This study is the first and intended to initiate more research in fitting distributions to weather indices. Our results can be used in stochastic risk management and in the design of insurance products [13, 14, 15].

This study presents models of Philippine precipitation using daily rainfall data from the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA). Specifically, we fulfill the following objectives: (i) Fit the daily and cumulative rainfall data into positively-skewed distributions; (ii) Determine the most appropriate probability distribution using goodness-of-fit tests; and (iii) Determine the subsets of the rainfall data where the fitted distributions exhibit invariance. The third objective is intended to justify forecasts that only use subset of the total rainfall data set. That is, whatever subset is used, a similar probability distribution can be invoked up to some degree. This is important especially when only a subset of the rainfall data is available.

In this study, we utilize rainfall data available from 28 synoptic stations of PAGASA, where each region of the Philippines is represented. We consider 55-year daily rainfall data from January 1961 to August 2016. Each station reaches area of 30 km up to 50 km radius. The daily rainfall data is positively-skewed where most data points fall under lower values and is heavily-tailed [7]. We use different probability distributions for fitting, e.g., Negative Binomial, Gamma, Exponential, Weibull, Poisson, Geometric and Normal. Our results show that Gamma distribution is a good fit for the daily up to the ten-day cumulative rainfall data. Since the model does not use time series analysis, we can only predict the rainfall amount but not when that event will exactly happen.

## 3 Related Studies

### 3.1 Foreign Studies

The following are some of the existing researches regarding the use of probability distributions to model rainfall data.

#### 3.1.1 Quarterly Data

Dikko et al. (2013) [16] developed a model from the quarterly rainfall data of Zaria, Nigeria using Gamma distribution. Using the Kolmogorov-Smirnov test, they concluded that Gamma model adequately fit the rainfall data.

#### 3.1.2 Monthly Data

In 1999, Sen et al. [17] modeled the monthly rainfall data in Libya. They utilized the latest recorded rainfall data for at least 20 years and confirmed that monthly rainfall is Gamma distributed using chi-square test.

A similar research was studied by Al-Suhili et al. [18] where they modeled the monthly rainfall data of Sulaimania Region in Iraq. They fitted continuous distributions such as, Normal, Lognormal, Weibull, Exponential and Gamma. From their calculations, they found that Gamma distribution best fit the data.

### 3.1.3 Daily Data

As early as in 1998, probability distributions were used to model rainfall. Duan et al. [19] worked on the Weibull distribution to model the daily precipitation in the US Pacific Northwest. They have seen that the one-parameter Weibull distribution performed as efficient compared with the Gamma model.

Li et al. (2012) [7] conducted studies on the 49 daily precipitation records in Texas. They worked on the Kappa, Gamma, Exponential and Generalized Pareto (GP) distributions. They also used hybrid (Gamma and GP) distributions to compensate for low-generating of extreme values of the first four distributions. They have shown that application of hybrid distributions can create an efficient and more realistic rainfall simulations.

In a study conducted by Brissette et al. (2012) [8], rainfall data were collected from 24 weather stations and two watersheds in Quebec, Canada. They evaluated six probability distributions, namely, Exponential, Gamma, Weibull, skewed-Normal, Mixed-Exponential, and a hybrid (exponential and GP). Their results showed that Mixed-Exponential distributions best simulates the daily precipitation in the province of Quebec.

Another study by Neykov et al. (2014) [20] modeled rainfall data using Gamma, Weibull and Lognormal distributions. Similar with Li et al. (2012), they also used hybrid distributions since common distribution models fail to produce sufficiently heavy-tail results. Based from their results, hybrid distributions best fit the daily rainfall data.

### 3.1.4 Hourly Data

Dan'azumi et al. (2010) examined hourly rainfall data in Peninsular Malaysia. They tried four distributions (GP, Exponential, Gamma and Beta). They evaluated the models using goodness-of-fit tests and concluded that Generalized Pareto distribution is the best model for hourly rainfall data [21].

## 3.2 Local Studies (Philippines)

Studies on rainfall patterns have been made. Pajuelas (2000) [10] researched on rainfall variations in the Philippines in 1950-1996. Long-term trends and variability of the rainfall in different parts of the country were also analyzed using time series analysis as in the research of Villafuerte et al. (2014) [5] and Cinco et al. (2014) [11]. June to October is the usual rainy season in the Philippines. However, there is still no existing published literature for modeling the daily rainfall data in the Philippines using probability distributions.

We have done a preliminary study to demonstrate ways to forecast rainfall [22]. These forecasts were used to predict the number of claims for the Weather-Index Based Insurance (WIBI) in Malaybalay City and Davao City. The Moving Averages technique was applied to 30-year rainfall data. It was observed that this technique did not produce a satisfying forecast since it did not capture the rare extreme cases valuable [23] for WIBI. Another approach used to forecast rainfall was fitting probability distributions to the rainfall data. We evaluated probability distributions (Exponential, Negative Binomial, Gamma, Poisson and Normal) using

Pearson's chi-square goodness-of-fit test. Results show that daily and cumulative rainfall data in Malaybalay City and Davao City follow the Gamma distribution. This can be used to reasonably estimate the number of WIBI claims.

## 4 Methodology

### 4.1 Data Collection

The 55-year daily rainfall data (January 1961 - August 2016) for 28 synoptic stations (as shown in Fig. 1) is requested from PAGASA. These data are examined for sufficiency and removal of missing values.

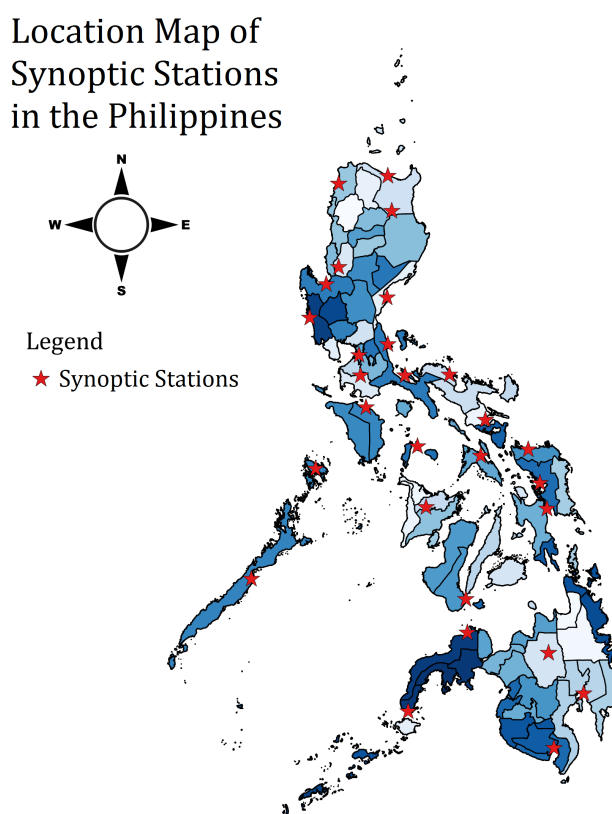


Figure 1: Philippine map presenting the locations of the 28 PAGASA weather stations

We construct the histograms using intervals of 10mm as bins to represent the classes of rainfall amounts. We then calculated the relative frequencies that will be used for the probability distribution fitting. The probability distributions are fitted for each rainfall data in each station, first by adjusting the 10 mm bins (e.g., if  $i = 10\text{mm}$  then  $i_{adj} = (i/10) + 0.5$ ). Then, these distributions are evaluated using goodness-of-fit tests.

We have tested the data set for invariance using two versions. *Invariance version 1*: The daily rainfall data is examined in exhibiting invariant patterns for different time scales (ten-year, five-year, one-year, half-year, and monthly). Using chi-square test, the most appropriate fit to each of these subsets is chosen. *Invariance version 2*: We have also tested if a good fit still remains if we use the distribution with parameter estimated using the complete data set as fit to the subsets.

If the rainfall data demonstrates invariance, then researchers may opt to request and use smaller sets of data without sacrificing its integrity of prediction. We can still produce substantial outcome necessary for our study since the data sets still follow the same distribution. It will also be more time and cost-efficient to deal with smaller sets of data.

Cumulative rainfall data (adding rainfall amount of  $n$  consecutive days, where  $n = 2, 3, \dots, 10$ ) are also fitted to the probability distributions. We also apply goodness-of-fit tests to determine the most fit distribution.

## 4.2 Probability Distributions

Daily rainfall data are usually positively skewed and heavy-tailed, that is, most data points fall on the side with the lower values [7]. Based from our initial findings using the daily rainfall data available for Malaybalay City and Davao City, a special probability distribution (i.e., Gamma) suitably fit the daily rainfall data. We focus on this distribution but we have also tested Negative Binomial, Exponential, Weibull, Poisson, Geometric and Normal distributions for fitting. Negative Binomial distribution is a good candidate for fitting daily rainfall data, but the Gamma distribution is good both for fitting daily and cumulative rainfall.

**Gamma Distribution.** Let  $x$  be the random variable for rainfall amount in mm. The Gamma distribution has two parameters, where  $\alpha$  is the shape parameter and  $\beta$  is the rate parameter. It has a probability density function given by:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, x \geq 0. \quad (1)$$

## 4.3 Goodness-of-fit Test

To assess the consistency of the probability distributions in modeling the daily rainfall data of the Philippines, goodness-of-fit tests are used. We use Pearson's chi-square test, and the maximum log-likelihood to estimate the Akaike Information Criterion (AIC). In this paper, we only present the commonly used Pearson's chi-square statistic. The Pearson's chi-square test will be used to determine if the daily rainfall data are consistent with the probability distribution. The null and alternative hypotheses to be tested are stated as follows:

$H_o$ : The rainfall data do follow the specified distribution.

$H_a$ : The rainfall data do not follow the specified distribution.

To compute for the  $\chi^2$  statistic, we employ the following formula:

$$\chi_c^2 = \sum_{i=1}^n \frac{(O_i - T_i)^2}{T_i} \quad (2)$$

where,  $\chi_c^2$  is the chi-square statistic,  $O_i$  is the observed value in bin  $i$  and,  $T_i$  is the expected value in bin  $i$ . We consider  $\alpha = 0.05$  as the level of significance and, the degrees of freedom ( $df$ ) is the total number of bins ( $n$ ) less by 1 and less the number of parameters of the distribution,  $k$  (i.e.,  $df = n - k - 1$ ). For instance, the  $df$  if we fit an exponential distribution is  $df = n - 2$ . We reject the null hypothesis  $H_o$  if  $\chi_{\alpha[df]}^2 < \chi_c^2$ . Otherwise, we fail to reject  $H_o$ .

The rule-of-thumb in using this test is that the expected frequency for all the bins must be  $\geq 5$  [24]. In this study, we are analyzing a large data set where extreme rare events (large rainfall amounts) are important. Majority of the bins have frequencies  $\geq 5$  but a few have frequencies  $< 5$  to reflect the rare extreme events.

## 4.4 Distribution Fitting

To determine which probability distribution is a good fit, we formulate a nonlinear program (NLP) that minimizes the  $\chi_c^2$  value. A sample NLP for fitting Gamma distribution is as follows:

*Minimize*

$$\chi_c^2 = \sum_{i=1}^n \frac{(O_i - T_i)^2}{T_i}$$

*subject to*

$$\begin{aligned} \sum_{i=1}^n T_i &\leq 1 \\ \alpha, \beta &> 0 \end{aligned}$$

where  $\alpha$  and  $\beta$  are the shape and rate parameters, respectively;  $O_i$  is the relative frequency of bin  $i$ ; and,  $T_i$  is the theoretical Gamma frequency of bin  $i$ .

A similar formulation is done for each station for all considered probability distributions. We solve the NLP above using MS Excel Solver with RGRNonlinear engine. Sample computations are presented in the Supplementary File.

## 5 Results and Discussion

### 5.1 Fitting to 1961-2016 Data

The histogram for the daily rainfall data for each station are constructed and fitted to the probability distributions considered in this study. As an example, the fitted distributions for Casiguran, Aurora station are given in Fig. 2. The chi-square value and parameter estimates for the respective distribution are shown in the figure.

Although selected probability distributions fit well on the daily rainfall data using the chi-square test, the Gamma distribution gives the lowest chi-square statistic. In fact, this is also the case for the other 27 stations which can be seen in Table 1. Since all chi-square values are less than 1, then we fail to reject the null hypothesis, that is, the rainfall data is consistent with the Gamma distribution. Some relevant statistics are also available in the said table.

One important characteristic of the distribution is the higher variance compared to the mean ( $\sigma^2 > \mu$ ). This implies heavy-tailed distribution. Moreover, we have observed a limitation of the fitted Gamma distribution. The theoretical mean based on the fitted Gamma distribution overestimates the sample mean (arithmetic mean of the data set). If this limitation will affect our decision-making (e.g., in calculating claims in insurance), the sample mean and sample variance can be used instead.

The chi-square test determines how good a fit is based from the total sum of squares of the deviations between the observed and expected frequencies [24]. In real-world scenario, the most valuable result of modeling rainfall is its ability to capture extreme cases such as scarcity and excessive rainfall [23].

Even if it is concluded that the Gamma distribution is the most appropriate fit in all stations, there are many instances where the parameters generated from the distribution do not adequately fit these extreme events. In particular, the 0mm rainfall bin is often underestimated like in Fig. 3. This underestimation of the probability for 0mm rainfall leads to the overestimation of the mean (theoretical vs sample mean). If our concern is the risk of heavy rainfall

Table 1: Chi-Square Statistic, Gamma Parameters, Mean and Variance of the Daily Rainfall

Stations	$\chi_c^2$ Statistic	$\alpha$	$\beta$	Sample $\mu$	Theoretical $\mu$	Sample $\sigma^2$	Theoretical $\sigma^2$
Alabat	0.4099	0.4845	5.3377	12.1131	21.8099	537.7867	1257.9252
Ambulong	0.3812	0.4674	3.8345	7.3937	14.1058	290.2251	608.2298
Aparri	0.3712	0.4808	3.6777	7.6806	13.8335	351.3177	575.6597
Baguio	0.4450	0.4055	6.6438	13.1523	23.1054	1251.4053	1618.0082
Calapan	0.2044	0.7279	2.2392	9.0945	11.9691	291.8176	332.6180
Casiguran	0.1997	0.5683	3.5382	13.2524	15.9722	689.2780	644.3421
Catarman	0.2640	0.6063	3.6925	13.1551	18.1223	547.6203	759.5473
Catbalogan	0.2779	0.7381	2.7476	11.3585	15.8309	365.2982	518.9384
Coron	0.1453	0.5623	2.5394	9.2205	10.3486	378.5312	317.2361
Daet	0.3016	0.5815	3.9648	12.9429	18.8248	565.5086	838.3145
Dagupan	0.3555	0.4616	3.6535	8.7230	13.0997	513.0897	541.9259
Davao	0.0921	0.9189	1.4883	8.0656	9.1687	184.2827	268.2910
Dipolog	0.1652	0.7159	2.2587	9.5762	11.8642	295.4741	331.9550
Dumaguete	0.1898	0.7041	1.9076	6.1234	9.2495	119.3114	227.2054
General Santos	0.1030	0.8198	1.3137	30.9043	32.1319	801.4443	785.0553
Iba	0.3619	0.4297	4.5369	12.2384	15.7475	932.5550	784.9772
Infanta	0.2191	0.7734	2.9654	15.0836	18.3845	614.8149	642.2330
Laoag	0.6282	0.4275	4.2408	7.4816	14.4275	624.4602	678.0803
Legaspi	0.3223	0.6283	3.7886	13.2480	19.4694	509.1944	834.9838
Malaybalay	0.1577	1.0125	1.8059	11.0668	13.5575	232.3778	500.6942
Masbate	0.4903	0.4904	4.6714	8.5683	18.8997	279.6118	968.3227
Puerto Princesa	0.1430	0.6564	1.9648	6.7124	8.8297	175.3931	221.7624
Romblon	0.1848	0.6560	2.4711	8.9604	12.0079	276.2044	360.5720
Roxas City	0.1836	0.6070	2.5494	8.3154	11.3981	261.5332	350.6694
Science Garden	0.2519	0.5000	3.5888	9.7572	14.0363	438.8623	572.2657
Tacloban	0.3194	0.7170	2.8582	10.6722	16.0729	293.7729	544.1465
Tuguegarao	0.3482	0.5176	3.1253	7.1080	12.2818	314.6867	445.7025
Zamboanga	0.1125	0.7481	1.5527	5.8815	7.4515	130.7624	157.4928

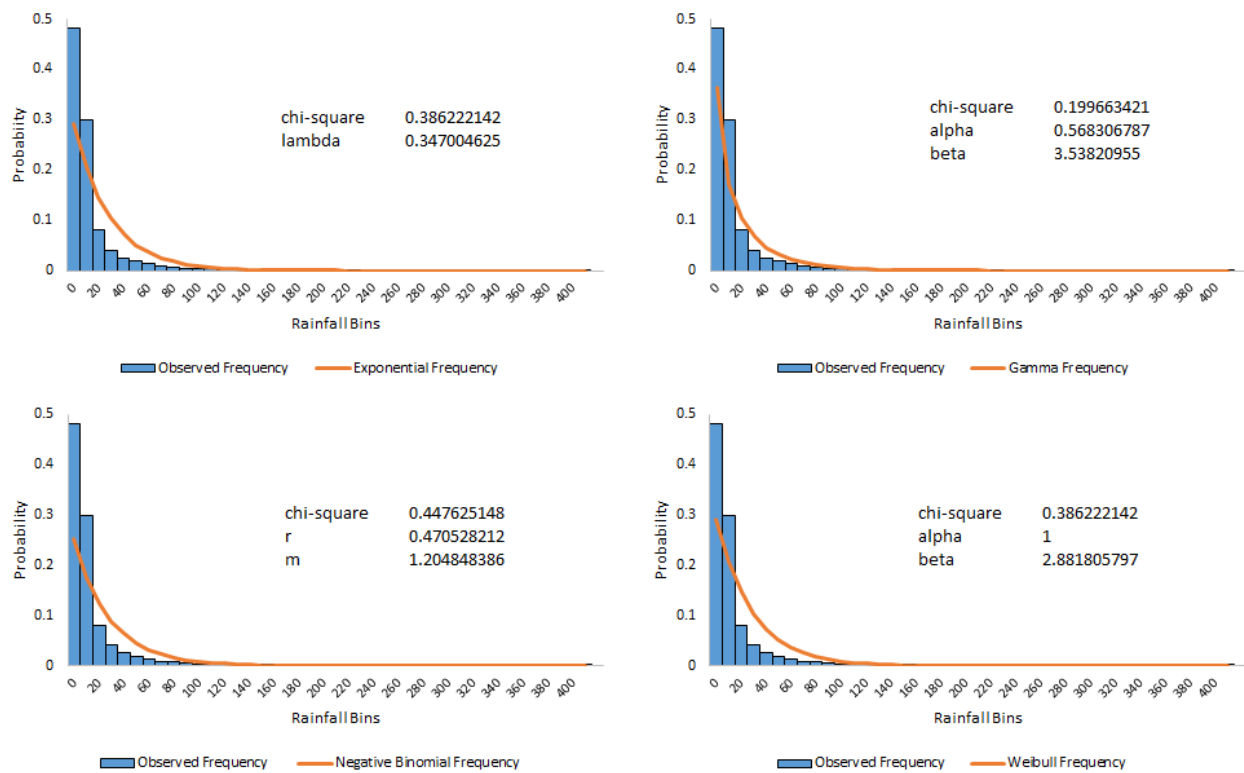


Figure 2: Rainfall histogram for Casiguran Station and selected fitted probability distributions with corresponding parameters

and not the risk of drought, the fitted Gamma distribution and its corresponding theoretical mean can be used. The difference between the risk cost in using the theoretical mean and the risk cost in using the sample mean can be part of the insurance safety loading.

## 5.2 Fitting to Subsets of the Data (Invariance Version 1)

We have also employed fitting smaller sets (ten-year, five-year, one-year, half-year, and monthly) of rainfall data into probability distributions. If we take a look at the chi-square values for the different subsets of rainfall data shown in Table 2, these are all less than 1. Meaning, these subsets of data considered are still consistent with the Gamma distribution.

In Fig. 4, we can visually compare the histograms for each data sets. Even if the chi-square test indicates that the Gamma distribution is a good fit for smaller sets of rainfall data, it is possible that the Gamma distribution for the monthly and half-year data would not produce sufficient rainfall amount prediction. This means that decreasing the quantity of data considered would only be suitable up to some extent.

## 5.3 Fitting the Parameters from 1961-2016 Data to Subsets of the Data (Invariance Version 2)

In this section, we fit the obtained Gamma parameters in Table 1 to different subsets of the rainfall data. As we can see in Table 3, most of the chi-square statistics are still within the acceptance level. That is to say that majority of the subsets of rainfall follow the same Gamma distribution as the 55-year daily rainfall data.



Table 2: Chi-Square Statistic for Different Time Scales of Data (Invariance Version 1)

Stations	Time Scales					
	1961-2016	10-year	5-year	1-year	Half-year	Monthly
Alabat	0.4099	0.1813	0.2269	0.2133	0.2449	0.4578
Ambulong	0.3812	0.2170	0.2212	0.2646	0.3593	0.4006
Aparri	0.3712	0.2951	0.3952	0.5565	0.5565	0.2196
Baguio	0.4450	0.4894	0.3921	0.6532	0.2306	0.6289
Calapan	0.2044	0.2910	0.2963	0.3748	0.1536	0.1536
Casiguran	0.1997	0.1880	0.2175	0.5219	0.4041	0.2297
Catarman	0.2640	0.2466	0.2569	0.5493	0.2459	0.2836
Catbalogan	0.2779	0.3448	0.3814	0.4406	0.2620	0.2103
Coron	0.1453	0.1506	0.2096	0.3888	0.6215	0.7415
Daet	0.3016	0.2167	0.2527	0.2635	0.2996	0.4880
Dagupan	0.3555	0.4891	0.5215	0.3575	0.5807	0.4540
Davao	0.0921	0.1343	0.1387	0.2144	0.1979	0.1315
Dipolog	0.1652	0.1504	0.1920	0.1221	0.4664	0.3827
Dumaguete	0.1898	0.2693	0.1277	0.0746	0.4673	0.3678
General Santos	0.1030	0.0899	0.0728	0.2915	0.1517	0.2066
Iba	0.3619	0.4092	0.5039	0.8116	0.8643	0.9299
Infanta	0.2191	0.2603	0.3275	0.7283	0.1047	0.2162
Laoag	0.6282	0.7257	0.6529	0.9957	0.9540	0.4915
Legaspi	0.3223	0.3707	0.2459	0.3990	0.2015	0.2678
Malaybalay	0.1577	0.2290	0.2464	0.1223	0.2227	0.1559
Masbate	0.4903	0.1647	0.1833	0.4396	0.2792	0.3348
Puerto Princesa	0.1430	0.0349	0.1541	0.2451	0.4020	0.0349
Romblon	0.1848	0.2163	0.1995	0.3703	0.1411	0.1796
Roxas City	0.1836	0.2407	0.2757	0.6046	0.0469	0.2678
Science Garden	0.2519	0.3236	0.3112	0.3909	0.5006	0.5523
Tacloban	0.3194	0.4086	0.4678	0.3201	0.3828	0.1647
Tuguegarao	0.3482	0.2841	0.3275	0.1974	0.4254	0.1352
Zamboanga	0.1125	0.1009	0.1391	0.2943	0.2381	0.3654

Table 3: Chi-Square Values for Invariance Version 2

Stations	Time Scales					
	1961-2016	10-year	5-year	1-year	Half-year	Monthly
Alabat	0.4099	0.3789	0.4001	0.4301	0.4268	0.4795
Ambulong	0.3812	0.3339	0.3043	0.4007	0.6522	0.9645
Aparri	0.3712	0.4043	0.4243	0.5902	0.5565	0.5902
Baguio	0.4450	0.5408	0.4179	0.8369	0.6106	1.5769
Calapan	0.2044	0.4010	0.4059	0.7633	0.2123	0.1526
Casiguran	0.1997	0.1934	0.2398	1.9348	0.4464	0.4324
Catarman	0.2640	0.2516	0.3010	0.6190	0.2587	0.6237
Catbalogan	0.2779	0.6701	1.7702	0.4496	0.3462	0.4261
Coron	0.1453	0.1522	0.2845	0.8500	0.6302	2.7835
Daet	0.3016	0.2516	0.2707	0.2958	0.3257	0.7008
Dagupan	0.3555	1.3207	1.3251	0.3639	0.6849	4.6998
Davao	0.0921	0.1550	0.1609	0.4285	0.2087	0.2306
Dipolog	0.1652	0.1523	0.2161	0.2010	0.8311	0.3871
Dumaguete	0.1898	0.8103	0.1471	0.1525	0.4756	0.3730
General Santos	0.1030	0.0920	0.0855	0.4425	0.2610	0.3101
Iba	0.3619	0.4461	0.8017	1.6427	1.1215	8.1793
Infanta	0.2191	0.2981	0.5690	10.1170	0.1181	0.2805
Laoag	0.6282	1.0139	0.6961	1.2925	0.9654	4.0310
Legaspi	0.3223	0.5218	0.2527	0.5784	0.2627	0.4826
Malaybalay	0.1577	0.6180	1.8447	0.1805	0.3204	0.2230
Masbate	0.4903	0.4488	0.4171	0.6938	0.5799	0.3790
Puerto Princesa	0.1430	0.1414	0.1618	0.2481	0.4046	5.9052
Romblon	0.1848	0.6716	0.2147	0.6637	0.1985	0.2007
Roxas City	0.1836	0.3366	0.4286	11.3534	0.3017	0.3615
Science Garden	0.2519	0.9628	0.6835	0.5306	0.5056	2.5424
Tacloban	0.3194	0.9920	2.8359	0.3301	0.5046	0.2035
Tuguegarao	0.3482	0.3131	0.3317	0.3439	0.5171	0.3042
Zamboanga	0.1125	0.1066	0.1512	0.7289	0.2536	0.7689

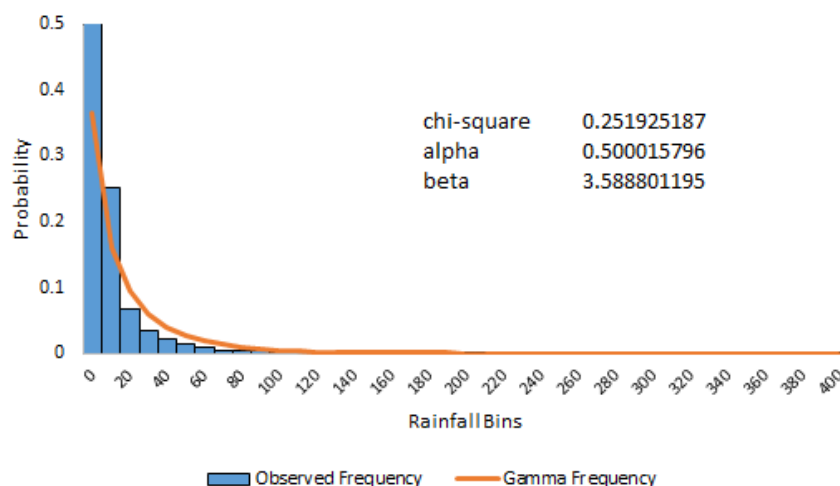


Figure 3: Gamma Distribution for the Daily Rainfall in Science Garden Quezon City Station

## 5.4 Fitting to Cumulative Data

Similar to the daily data, we have also fitted the probability distributions into the cumulative rainfall data. Using the chi-square test, the cumulative  $n$ -day rainfall patterns where  $n = 2, 3, \dots, 10$  also follow Gamma distribution. It is evident that the parameters  $\alpha$  and  $\beta$  increase as  $n$  increases. That is because the probability of a zero rainfall decreases as we increase the number of days considered. In some stations, there is an obvious change in the shape when  $\alpha < 1$  to  $\alpha > 1$  as we can see in Fig. 5.

## 6 Conclusions

Considering the 55-year daily rainfall data available from 28 different synoptic stations in the Philippines with each region represented, the rainfall distribution for each station follows the Gamma distribution. With regards to the cumulative rainfall amount, we have confirmed that each stations rainfall pattern still follows the Gamma distribution. Lastly, we considered modeling the rainfall pattern using smaller sets of the data and have shown that still, it is consistent with the Gamma distribution.

The use of relatively simple probability distributions in this research had generated novel results to represent the Philippine rainfall data. However, we recommend the following areas to work on in the future: (i) Since the distributions generated are for areas covered by the 28 weather station, we can consider spatial interpolation to generate a distribution for the uncovered areas; and (ii) As mentioned earlier, there are stations where the resulting distribution for lower bins are underestimated. We may opt to use hybrid distributions to improve the probability estimates.

### Supplementary files:

- Summary of Gamma distribution parameters.xlsx
- Sample computations.xlsx

**Acknowledgments:** JFR is supported by the UP System Enhanced Creative Work and Research Grant (ECWRG 2016-1-009).

**Author Contributions:** JFR conceived the study; DCNC provided the input information; VABB and ABBA produced the results; all authors contributed in the analysis, interpretation,

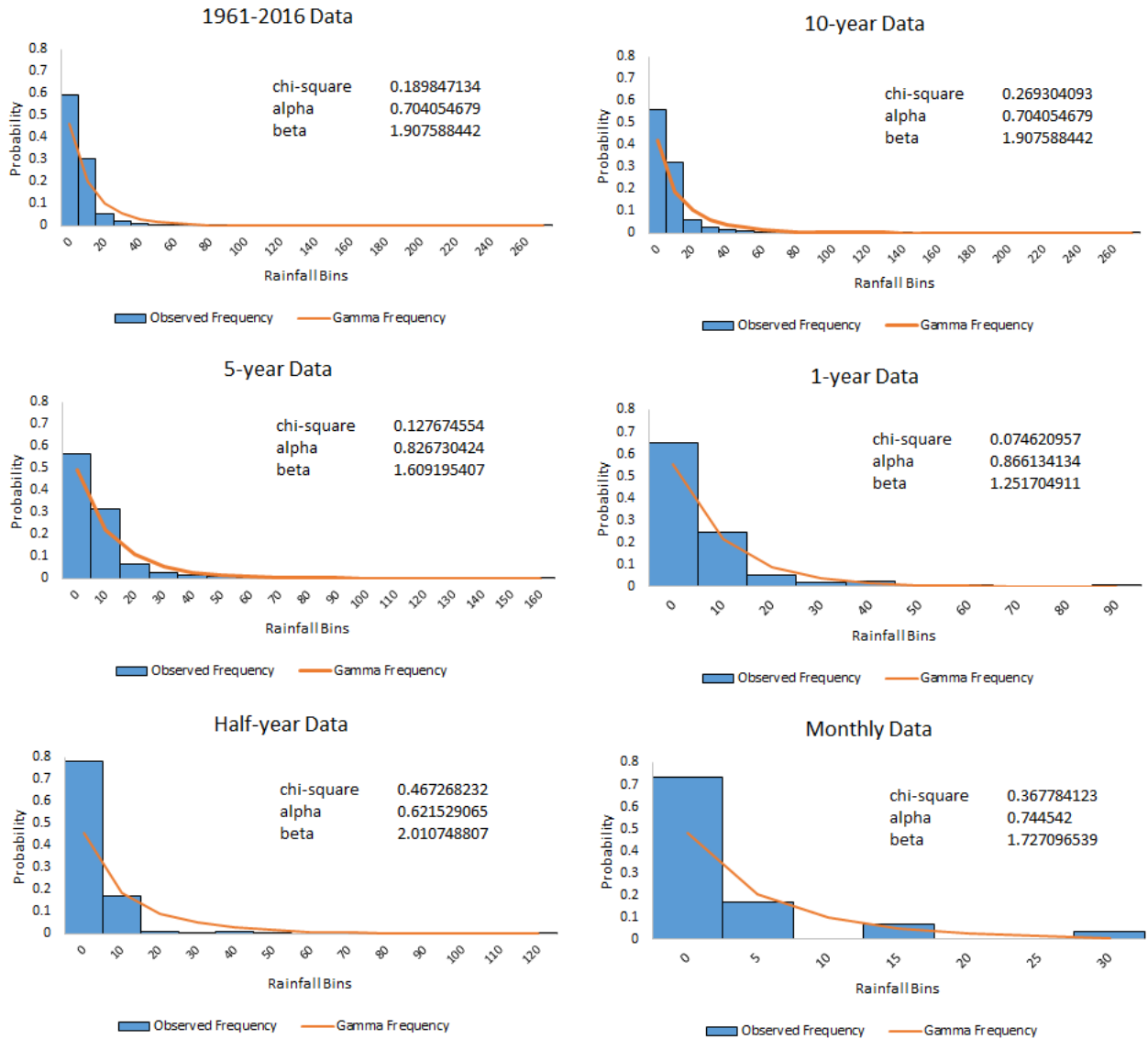


Figure 4: Gamma Distribution for the Daily Rainfall in Dumaguete Station using Different Time Scales

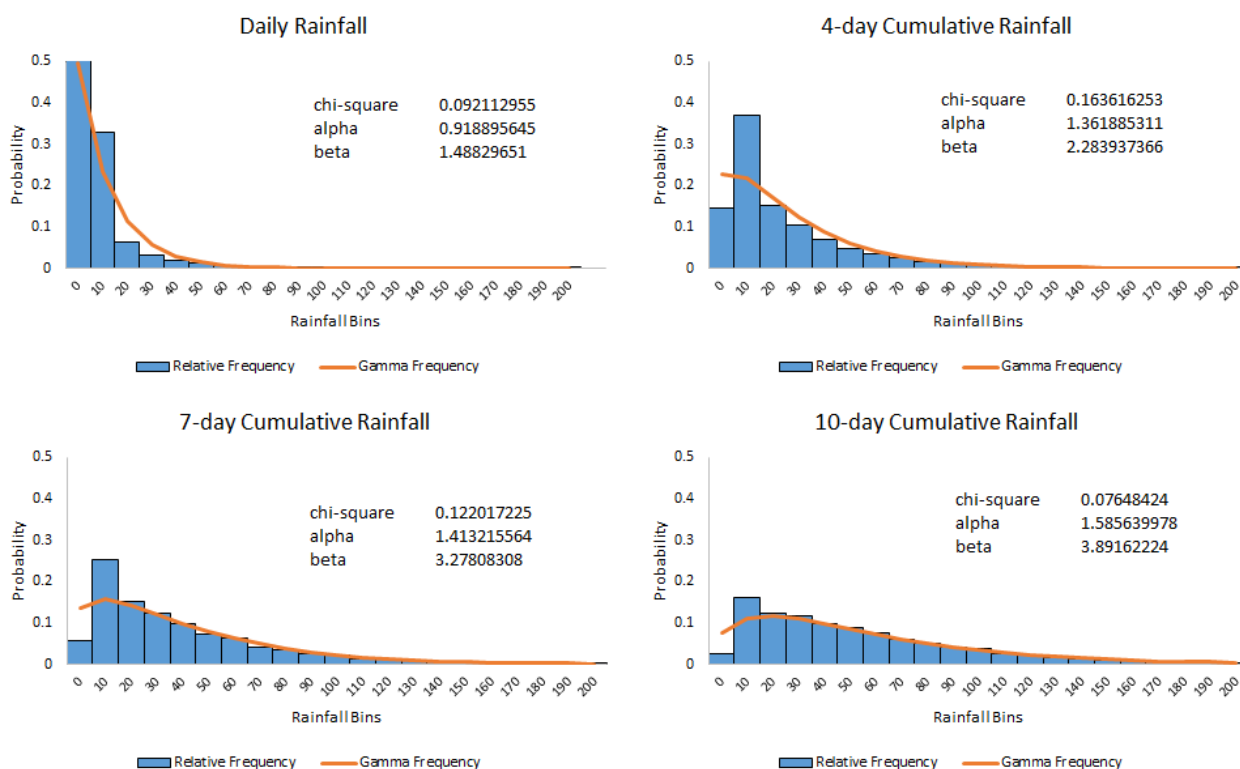


Figure 5: Gamma Distribution for the Cumulative Rainfall in Davao City Station

discussion and presentation of the results; VABB and JFR wrote the manuscript with VABB as the lead author; all authors reviewed the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

- [1] Philippine Statistics Authority, "Countrystat Philippines," 2014. Accessed: May 13, 2017.
- [2] F. P. Lansigan, "Climate risk for weather index-based insurance for crop production in the Philippines," 2013.
- [3] T. A. Cinco, R. G. de Guzman, A. M. D. Ortiz, R. J. P. Delfino, R. D. Lasco, F. D. Hilario, E. L. Juanillo, R. Barba, and E. D. Ares, "Observed trends and impacts of tropical cyclones in the Philippines," *International Journal of Climatology*, vol. 36, pp. 4638–4650, 2016.
- [4] R. N. Concepcion, M. V. O. Espaldon, S. Lewicki-Dhainaut, R. M. M. Dedicataria, and E. R. Abuca, "Coping with extreme climatic events: Stories of resiliency in the Philippines," *KC3 SEARCA*, vol. 1, no. 4, pp. 1–2, 2011.
- [5] M. Q. I. Villafuerte, J. Matsumoto, I. Akasaka, H. G. Takahashi, H. Kubota, and T. A. Cinco, "Long-term trends and variability of rainfall extremes in the Philippines," *Atmospheric Research*, vol. 137, pp. 1–13, 2014.
- [6] H. Aksoy, "Use of gamma distribution in hydrological analysis," *Turkish Journal of Engineering and Environmental Sciences*, vol. 24, pp. 419–428, 2000.

- [7] C. Li, V. P. Singh, and A. K. Mishra, "Simulation of the entire range of daily precipitation using a hybrid probability distribution," *Water Resources Research*, vol. 48, p. W03521, 2012.
- [8] Z. Li, F. Brissette, and J. Chen, "Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds," *Hydrological Processes*, 2012.
- [9] S. W. Martin, B. J. Barnett, and K. H. Coble, "Developing and pricing precipitation insurance," *Journal of Agricultural and Resource Economics*, vol. 26, pp. 261–274, 2001.
- [10] B. G. Pajuelas, "A study of rainfall variations in the Philippines: 1950-1996," *Science Diliman*, vol. 12, no. 1, pp. 1–28, 2000.
- [11] T. A. Cinco, R. G. de Guzman, F. D. Hilario, and D. M. Wilson, "Long-term trends and extremes in observed daily precipitation and near surface air temperature in the Philippines for the period 1951-2010," *Atmospheric Research*, vol. 145146, p. 1226, 2014.
- [12] I. Akasaka, "Interannual variations in seasonal march of rainfall in the Philippines," *International Journal of Climatology*, vol. 30, pp. 1301–1314, 2010.
- [13] J. D. Daron and D. A. Stainforth, "Assessing pricing assumptions for weather index insurance in a changing climate," *Climate Risk Management*, vol. 1, pp. 76–91, 2014.
- [14] S. Conradt, R. Finger, and M. Sporri, "Flexible weather index-based insurance design," *Climate Risk Management*, vol. 10, pp. 106–117, 2015.
- [15] A. D. Zapranis and A. A. K., *Weather Derivatives: Modeling and Pricing Weather-Related Risk*. Springer, 2013.
- [16] H. G. Dikko, I. J. David, and H. Bakari, "Modeling the distribution of rainfall intensity using quarterly data," *IOSR Journal of Mathematics*, vol. 9, no. 1, pp. 11–16, 2013.
- [17] Z. Sen and A. G. Eljadid, "Rainfall distribution function for Libya and rainfall prediction," *Hydrological Sciences Journal*, vol. 44, no. 5, pp. 665–680, 1999.
- [18] R. H. Al-Suhili and R. Khanbilvardi, "Frequency analysis of the monthly rainfall data at Sulaimania region, Iraq," *American Journal of Engineering Research*, vol. 3, no. 5, pp. 212–222, 2013.
- [19] J. Duan, J. Belker, and G. E. Grant, "Evaluation of probability density functions in precipitation models for the Pacific Northwest," *Journal of the American Water Resources Association*, vol. 34, no. 3, pp. 617–627, 1998.
- [20] N. M. Neykov, P. N. Neytchev, and W. Zucchini, "Stochastic daily precipitation model with a heavy-tailed component," *Natural Hazards and Earth System Sciences*, vol. 14, pp. 2321–2335, 2014.
- [21] S. Danazumi, S. Shamsudin, and A. Aris, "Modeling the distribution of rainfall intensity using hourly data," *American Journal of Environmental Sciences*, vol. 6, no. 3, pp. 238–243, 2010.
- [22] A. B. B. Abilgos and J. F. Rabajante, "Forecasting of 30-year rainfall data in Davao city and Malaybalay city using moving averages technique and distribution fitting." Thesis (unpublished), 2016.

- [23] X. Zhang, L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, "Indices for monitoring changes in extremes based on daily temperature and precipitation data," *WIREs Climate Change*, 2011.
- [24] H. A. Taha, *Operations Research: An Introduction*. Prentice Hall, 8th ed., 2006.