Journal of Science Research (2010) Vol. 9: 69-73 *ISSN 11179333*

The Use of Gamma and Weibull Distributions in Modeling Rainfall Data in Nigeria (A comparative analysis)

Udomboso C. G., Chukwu A. U., and Nja M. E.

Abstract The distribution of rainfall in Nigeria is not uniform due to the slight differences in the climatic conditions from one geographical region to the other. The climatic conditions ranges from the 'very wet' mangrove forest zone of the coastal areas, especially the South-South, to the semi arid regions of the North-West and North-East that share boundaries with the 'very hot' desert zone of the North Africa subcontinent. Rainfall data are examples of environmental data, which generally can be modeled by the family of exponential distributions. The Weibull probability function is the most widely used in fitting the distribution of rainfall. This study compares the results obtained by this function with another distribution proposed to the African scientists, that is, the gamma probability function by employing the Kolmogorov-Smirnov (K-S) one sample test in testing the goodness-of-fit of these distributions. The results provided are useful tools for decision makers in hydrological and related establishments.

Keywords: Rainfall, Exponential family, Gamma distribution, Weibull distribution, K-S test.

Introduction

In hydrolo-gy and meteorology operations, the major challenge has been in forecasting rainfall events (heavy or spatial), and the area of greatest threat. Modeling rainfall provides forecasters with supplementary information on precipitation thresholds which can lead to significant flash or major flooding. The typical approach to gaining a better understanding of the spatial and temporal variability in precipitation starts with the acquisition of historical rainfall data [1]. Such historical data are invaluable in that they present necessary and sufficient information concerning accumulation amounts in both time and space and form the basis for fitting and testing distribution models. In cases where historical data is unavailable, or inaccurate or incomplete, geophysical models are resorted to in order to 'fill in' the missing values. These geophysical models are based on available data at other locations and times, as well as additional variables that add information to the model.

It is widely known that empirical distribution is inferior to parametric distribution model which allows for a more stable and extensive analysis, of which rainfall data is one. However for a distribution to be used, it must be flexible enough to represent a variety of rainfall regimes. The gamma and Weibull distributions are the most widely used of these distributions. This study notes here that rainfall distributions generally follow the family of exponential distributions just like any other environmental data. In this work, focus shall be on comparison of the results of the gamma and Weibull distributions. The estimation of the parameters of these distributions can be used to describe rainfall regimes and be used in various applications.

Maddox et al [2] studied the synoptic and meso-α aspects of more than 150 flash flooding events across the US. They used the storm data reports for the years 1933 – 1977 to compile a flood climatology which supplied essential information to identify a vast sample of intense precipitation events. Later Grice and Maddox [3] focused on heavy rain events in South Texas and the Texas Hill Country of the US, and defined an event by rainfall equal to or greater than 5 inches within 24 hours for the area excluding the Hill Country, and equal to or greater than 4 inches in 24 hours in the vicinity of the Hill Country.

Asakereh [1] noted that in many studies of climate change, assumption has often been made that only the mean, the location parameter, could change, while the shape of the distribution does not change. However he affirmed that the frequency of climate variable would be changed with change of location. Other researchers showed that the relative frequency of events depends on changes in the standard deviation [6, 7, 9]. This underscores the importance of using probability distribution in evaluating climate change (as is in the case of rainfall precipitation).

Udomboso C. G., Chukwu A. U. and Nja M.E.
Department of Statistics, University of Ibadan, Ibadan, Nigeria.

Results and Discussion

The EasyFit 5.0 was used in the analysis of the data which consisted of 1000 non-zero rainfall values collected for about 30 years in twenty (20) meteorological stations across the country; Maiduguri, Kano, Abuja, Yola, Jos, Bauchi, Ibadan, Osogbo, Benin, Owerri, Lokoja, Nguru, Warri, Yelwa, Kaduna, Calabar, Sokoto, Port Harcourt, Ikeja and Ilorin. We note here that rainfall pattern in Nigeria are similar (only that the duration of precipitation is observed to start earlier and ends later in the south, and gradually reduces towards the north).

In interpreting the significance of parameter estimates, we shall use the 2D-graphic presentation in figure 3 in describing the parameter space.

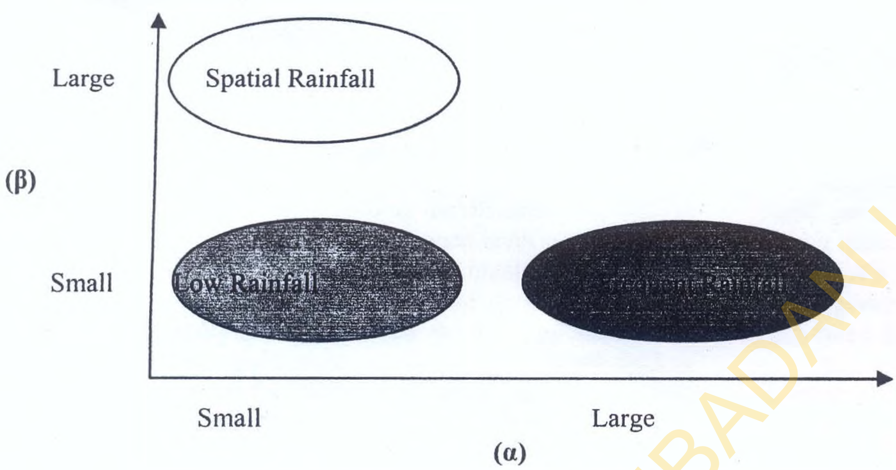


Figure 3 Rainfall Distribution based on the Parameter Space using the Gamma and Weibull Distributions

The figure summarizes that increase in the scale parameter results in infrequent rainfall and wide variance, and increase in the shape parameter results in high rainfall with tight variance. Thus a shape-dominated regime depicts a rainfall pattern that tends to be symmetrically distributed, indicating that drier-than-average events are as common as wetter-than-average events. The situation here is described as areas that typically receive consistent rainfall accumulations in the historical record. If the mean rainfall is constant, large value of $\hat{\alpha}$ will produce small values of $\hat{\beta}$. This shows a less variance in the distribution function. In the same vein, a scale-dominated regime depicts locations with large variance in comparison with the mean. Holding the mean rainfall constant, and increasing the $\hat{\beta}$ will reduce $\hat{\alpha}$, thus giving a more positively skewed distribution function.

We note here that both the shape and scale parameters must be involved in interpretation for meaningful prediction.

For the 2-parameter gamma distribution, $\hat{\alpha} = 0.80668$ and $\hat{\beta} = 14.025$. This makes the mean rainfall to have the value 11.3, and variance of 158.7.

For the 2-parameter Weibull distribution, $\hat{\alpha} = 0.86008$ and $\hat{\beta} = 10.451$. The mean rainfall events have the value 10.5, while the variance is 218.5.

This research is concerned with establishing a relationship between the estimated distribution parameters and the occurrence of extreme events. On a frank note, areas with scale-dominated rainfall most often would experience more extreme and abnormal events. This indicates that such areas would experience quite a range of rainfall amounts, so that there need be put in place infrastructure and plans to cope with extremely dry or wet conditions. Conversely, places having shape-dominated rainfall might experience more rain which may also result into large absolute variance. The large shape value, though, indicates a relatively consistent accumulation as the years go by.

Implementation and Results of the K-S One Sample Test

The goodness of fit (GOF) test measures the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the distribution selected fits to data.

The Kolmogorov-Smirnov test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function (ECDF). Assume that we have a random sample $x_1, \ldots, x_n$ from some distribution with CDF *F(x)*. The empirical CDF is denoted by

$$F_n(x) = \frac{1}{n}.[Number\ of\ Observations \geq n]$$

The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function:

$$D = \max_{1 \geq i \geq n}[F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i)]$$

The null and the alternative hypotheses are:

H_0: the data follow the specified distribution;

H_A: the data do not follow the specified distribution.

The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic, D, is greater than the critical value obtained from a table. The fixed values of α (0.01, 0.05 etc.) are generally used to evaluate the null hypothesis (H_0) at various significance levels. A value of 0.05 is typically used for most applications, however, in some critical industries, a lower α value may be applied.

The standard tables of critical values used for this test are only valid when testing whether a data set is from a completely specified distribution. If one or more distribution parameters are estimated, the results will be conservative: the actual significance level will be smaller than that given by the standard tables, and the probability that the fit will be rejected in error will be lower.

The P-value, in contrast to fixed α values, is calculated based on the test statistic, and denotes the threshold value of the significance level in the sense that the null hypothesis (H_0) will be accepted for all values of α less than the P-value. For example, if P=0.025, the null hypothesis will be accepted at all significance levels less than P (i.e. 0.01 and 0.02), and rejected at higher levels, including 0.05 and 0.1.

The P-value can be useful, in particular, when the null hypothesis is rejected at all predefined significance levels, and you need to know at which level it *could* be accepted.

EasyFit displays the P-values based on the Kolmogorov-Smirnov test statistics (D) calculated for each fitted distribution

The results of the Kolmogorov-Smirnov test are as follows:

Table 1 Goodness of Fit - Details

Gamma					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	1000 0.04301 0.04803 8				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.03393	0.03867	0.04294	0.048	0.05151
Reject?	Yes	Yes	Yes	No	No

Table 2 Goodness of Fit -Details

Weibull					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	1000 0.03402 0.19286 3				
□	0.2	0.1	0.05	0.02	0.01
Critical Value	0.03393	0.03867	0.04294	0. 048	0.05151
Reject?	Yes	No	No	No	No

While both distributions are good fits to the rainfall data, in comparison we note that the Weibull distribution performs better than the Gamma distribution. The Weibull, at significance level 0.1 and below, accept the null hypothesis, in contrast to the Gamma which accepts at alpha level 0.02 and below.

Industrial Applications

This study has many industrial applications, such as agriculture, transport, engineering, and so on. For instance, the parameters can provide useful insight into the suitability of an area in the provision of adequate rainfall for certain crops. Thus, stakeholders in agriculture may make decisions regarding crop health when considering the probability of no rainfall and dry events. This decision is important as it pose primary threats to productive agriculture. These predictions are also very useful to transporters (especially road and rail) in determining the transport systems, as extreme wet events may disrupt mobility. Dam and hydroelectric engineers and workers may be concerned with both wet and dry events in order to prevent the waterways from overflow, as well as maintaining the reservoir to last the entire dry season. Thus the definition of 'extreme' is different for each of these industries. In general, the parameters could go a long way in determining the probability of catastrophe related to extreme events.

References

[1] Asakereh, H. (2008). The Changes of Precipitation Frequency Distribution in Iran. *Journal of Humanities of The University of Isfahan* 30 (2), 81–90.

[2] Grice, G. K. and Maddox, R. A. (1982). Synoptic Aspects of Heavy Rain Events in South Texas Associated with the Westerlies. NOAA Tech. Memo. NWS SR – 106, National Weather Service Southern Region, Ft. Worth, TX, 21 pp.

[3] Husak, G. J., Michaelsen, J. and Funk, C. (2006). Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications. International Journal of Climatology. Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/joc.1441. Copyright of the Royal Meteorological Society.

[4] Ison, N. T., Feyerherm, A. M. and Dean, B. L. (1971). Wet Period Precipitation and the Gamma Distribution. *Journal of Applied Meteorology* 10, 658–665.

[5] Juras, J. (1994). Some Common Features of Probability Distributions for Precipitation. *Theoretical and Applied Climatology* 49, 69–76.

[6] Katz, R. W. (1991) Towards a Statistical Paradigm for Climate Change. 7th Conference on Applied Climatology, A nerican Meteorological Society, Boston.

[7] Katz, R. W. and Brown, B. G. (1992). Extreme Events in a Changing Climate : Variability is More Important than Averages. *Climate Change* 21(3), 289–302.

[8] Maddox, R. A., Chappell, C. F. and Hoxit, L. R. (1979). Synoptic and Meso- Scale Aspects of Flash Flood Events. *Bull. Amer. Meteor. Soc.* 60, 115–123.

[9] Mearns, L. O., Katz R. W. and Schneider, S. H. (1984). Extreme High-Temperature Events: Changes in their Probabilities with Changes in Mean Temperature. *J. Climate Appl. Meter.* 23, 1601–1613.

[10]1Ropelewski, C. F. and Halpert, M. C. (1996). Quantifying Southern Oscillation – Precipitation Relationships. *J. Climate* 9, 1043–1059.

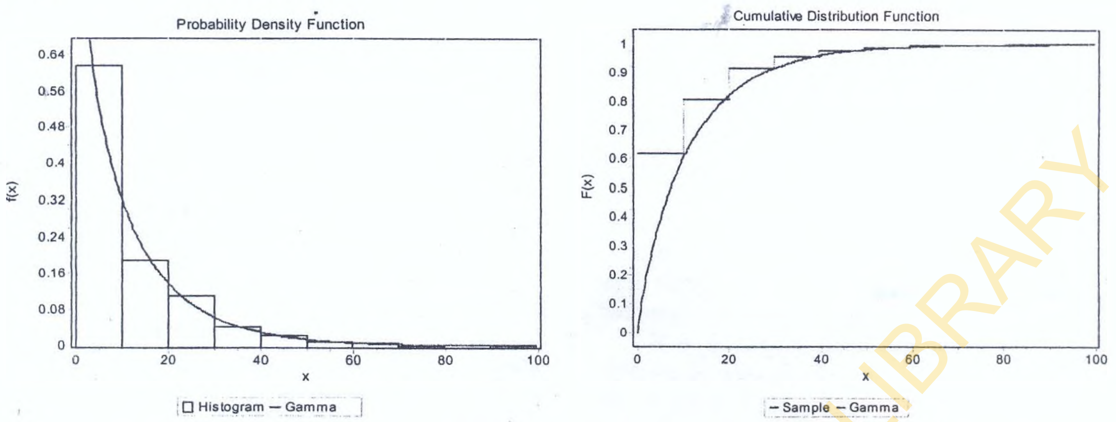
[11] Vinnikov, K. Y., Groisman, P. Y. and Lugina, K. M. (1990). Empirical Data on Contemporary Global Climate Changes (Temperature and Precipitation). J. Climate. 3, 662-677.

[12] Wilks, D. S. and Eggleston, K. L. (1992). Estimating Monthly and Seasonal Precipitation Distributions Using the 30- and 90-Day Outlooks. *J. Climate* 5(3), 252-259.

Appendix

Graphs of the Gamma and Weibull functions (*p.d.f* and *c.d.f*) based on the data used in the analysis are given below;

Gamma



Weibull