The Probability Distribution of Daily Precipitation at the Point and Catchment Scales in the United States

Lei Ye[1*], Lars S. Hanson[2], Pengqi Ding[1], Dingbao Wang[3], Richard M.Vogel[4]

1 School of Hydraulic Engineering, Dalian University of Technology, Dalian, China
2 Institute for Public Research, Center for Naval Analyses, Arlington, Virginia, USA.
3 Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, Florida, USA
4 Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts, USA

Abstract: Choosing a probability distribution to represent daily precipitation depths is important for precipitation frequency analysis, stochastic precipitation modeling and in climate trend assessments. Early studies identified the 2-parameter Gamma (G2) distribution as a suitable distribution for wet-day precipitation based on traditional goodness of fit tests. Here, probability plot correlation coefficients and L-moment diagrams are used to examine distributional alternatives for the wet-day series of daily precipitation for hundreds of stations at the point and catchment scales in the United States. Importantly, the G2 distribution performs poorly in comparison to either the Pearson Type-III (P3) or Kappa (KAP) distributions particularly for point rainfall. Our analysis indicates that the KAP distribution best describes the distribution of wet-day precipitation at the point scale, whereas the performance of G2 and P3 distributions are comparable for wet-day precipitation at the catchment scale, with P3 generally providing improved goodness of fit over G2. Since the G2 distribution is currently the most widely used probability density function, our findings could be considerably important, especially within the context of climate change investigations.

Key Words: Climate; Rainfall; Weather; L-moment diagram; PPCC; Pearson type III; Kappa; Gamma; Wet-day; Frequency analysis; Trend detection; Stochastic weather models

1. Introduction

Precipitation is paramount in the fields of hydrology, meteorology, climatology, and others. However, long series of precipitation data are not always available; therefore, establishing a probability distribution that provides a good fit to daily precipitation depths has long been a topic interest. Investigations into the probability distribution of daily precipitation can be found in at least three main research areas, namely, (1) stochastic precipitation models, (2) frequency analysis of precipitation, and (3) precipitation trends related to global climate change. Table 1 displays a sampling of the literature related to those three topics including the particular precipitation series and durations under investigation, and the proposed probability distributions recommended. Table 1 is by no means exhaustive; it only attempts to document the widespread interest in the

[*] Corresponding author. E-mail address: yelei@dlut.edu.cn

determination of a suitable distribution for daily precipitation totals in a wide range of studies across a wide range of fields of inquiry.

[*Table* 1 *goes here*]

1.1 Stochastic precipitation models:

Our central goal is to select a suitable generalized probability distribution for modeling daily precipitation depths, thus we are only concerned with the class of "two-part" stochastic daily precipitation models that utilize a probability distribution function to describe precipitation amounts on wet-days, while a probabilistic representation of precipitation occurrences can be separately described using a Markov model or some form of a stochastic renewal process (Buishand, 1978;Geng et al., 1986;Waymire and Gupta, 1981;Watterson, 2005). We only consider the selection of a suitable distribution for modeling wet-day daily rainfall, leaving the stochastic representation of the occurrence of zeros, to others.

It is evident from Table 1 that the wet-day precipitation series is the primary series considered within the stochastic precipitation model literature. Thom's (1951) suggestion of the 2-parameter Gamma (G2) distribution function for wet-day amounts seems to carry considerable weight. Buishand (1978) lent support to the suggestion of the G2 distribution by showing that for the wet-day series at six stations, the empirical Coefficient of Variation to Coefficient of Skewness ratio was quite close to the theoretical value of two for a G2 distribution. Geng et al. (1986) provided a review of other literature supporting the use of the G2 distribution for modeling wet-day rainfall.

While the G2 distribution is by far the most commonly advocated distribution for wet-day precipitation amounts, other distributions have also been suggested. Woolhiser and Roldan (1982), Wilks (1998) and Li et al. (2013) suggested the use of a three-parameter mixed exponential distribution instead of G2. Through a variety of goodness of fit tests and log-likelihood analyses, the mixed exponential was preferred to G2 (Wilks, 1998).

The Weibull (W2) and to a lesser extent the exponential distribution have also been suggested for modeling daily precipitation amounts (Duan et al., 1995;Burgueno et al., 2005). Duan et al. (1995) used a Chi-squared test to demonstrate that synthetic rainfall generated from the W2 and G2 models best match the observed daily rainfall data within each month. Burgueno et al. (2005) used graphical methods and the Kolmogorov-Smirnov test to give support to the W2 and exponential distributions.

1.2 Precipitation frequency analysis:

The second section of Table 1 displays a small portion of the literature related to precipitation frequency analyses. Since extreme rainfall values are of primary importance in these studies, censored series of rainfall (e.g. the Annual Maximum Series (AMS) and Partial Duration Series (PDS)) are often useful in these analyses (Stedinger et al., 1993). Table 1 displays that many of the precipitation frequency investigations of daily precipitation depths have selected the AMS series.

For many years, the most common approach to summarizing precipitation frequency analyses in the United States was the work of Hershfield (1961), which is commonly referred to as TP-40. Hershfield (1961) fitted a Gumbel distribution to the AMS of 24-hour precipitation. In the context of a national revision to the TP-40, Bonnin et al., (2006) fitted a generalized extreme value (GEV) distribution to the AMS of rainfall.

While the results of Bonnin et al. (2006) apply to the United States, other researchers have found similar results using similar methods in other parts of the world. Pilon et al. (1991) used L-moment goodness-of-fit results to show that the Gumbel distribution should be rejected in the favor of the GEV in Ontario, Canada. In Korea, Park and Jung (2002) successfully used the Kappa distribution (of which the GEV is a special case) to generate extreme precipitation quantile maps. In perhaps the most comprehensive assessment of the distribution of precipitation extremes, Papalexiou and Koutsoyiannis (2013) examined the goodness-of-fit of the GEV distribution to a global dataset of AMS. Analysis of such a large dataset enabled them to conclude that GEV models of AMS of daily precipitation provide a good approximation.

Interestingly, while a great deal of attention is given to fitting distributions to the relatively short AMS series of precipitation depths, very few studies directly explore the probability distribution of the complete series of daily precipitation (including zeros) or the wet-day series of daily precipitation (zeros excluded). Shoji and Kitaura (2006) investigated both complete and wet-day daily precipitation series, but included only the normal, lognormal, exponential, and W2 distributions as candidate distributions, and did not employ modern regional hydrologic methods such as the method of L-moments. Deidda and Puliga (2006) investigated the degree of left-censoring of wet-day series needed to fit a Generalized Pareto (GPA) distribution for 200 stations in Italy with a range of modern statistical analysis techniques. Wilson and Toumi (2005) derived a fundamental distribution for heavy rainfall, with a simple expression for rainfall as the product of mass flux, specific humidity and precipitation efficiency. Statistical theory predicted that the tail of the derived rainfall distribution has a stretched exponential form with a shape parameter of 2/3, which was verified by a global daily precipitation data set.

Perhaps the most thorough investigations, to date, on the probability distribution of daily precipitation amounts are the global studies by Papalexiou and Koutsoyiannis (2012, 2016). Papalexiou and Koutsoyiannis (2012) derived a generalized Gamma distribution (GG) from Entropy theory, using plausible constraints for wet-day series of daily precipitation series. Together, the two studies by Papalexiou and Koutsoyiannis (2012, 2016) revealed that the GG distribution provides a good approximation to the behavior of observed L-moments of global series of wet-day daily precipitation at 11,519 and 14,157 stations, respectively.

1.3 Precipitation trends and changes:

The third section of Table 1 summarizes a small portion of the precipitation trend literature which has become a rather large area of inquiry due to concerns over climate change, as evidenced from recent reviews on the subject (Easterling et al., 2000; Trenberth, 2011; Madsen et al., 2014). Almost universally, the G2 distribution appears to be accepted without serious consideration of alternative distributions. For instance, Groisman et al. (1999) compared maps of the empirical probability of summer 1-day rainfall exceeding 50.4 mm with maps of probabilities determined by a stochastic model using the fitted G2 distribution for the amounts. They found acceptable fits in regions where there are enough observed daily rainfall events greater than 50.4 mm.

This is an interesting contrast to the precipitation frequency analysis literature where a G2 distribution is often fit to wet-day series for the purpose of examining extreme rainfalls instead of using the AMS series fitted by a GEV or other distribution. Yoo et al. (2005) explained that conventional frequency analysis (using AMS) cannot expect to predict precipitation changes resulting from climate change; while an examination of the differences in the G2 distribution's parameters (fitted to the whole wet-day record) might predict such changes. They found that modifying the parameters of the daily G2 distribution can explain changes in rainfall quantiles predicted by General Circulation Models under various climate change scenarios.

In a national study of precipitation trends, Karl and Knight (1998) employed the G2 distribution to fill in missing precipitation observations. Both Watterson and Dix (2003) and Watterson (2005) assumed a G2 distribution for daily precipitation in the development of stochastic rainfall models for use in evaluating changes in precipitation extremes.

1.4 Research objectives:

In summary, there are a wide variety of previous studies which have explored the probability distribution of daily precipitation for the purposes of precipitation frequency analysis, stochastic precipitation modeling and for trend detection. There seems to be a consensus that annual maxima appear to be well approximated by either a GEV, Gumbel or GPA probability density function (pdf) and that series of wet-day daily precipitation totals are well approximated by a G2, GG, W2 or in some cases a mixed exponential pdf. However, other than the two recent global studies by Papalexiou and Koutsoyiannis (2012, 2016), we are unaware of any studies that have used recent developments in regional hydrologic frequency analysis such as L-moment diagrams or probability plot goodness of fit evaluations to evaluate the probability distribution of very large regional datasets comprised of the wet-day series of daily precipitation.

The recent studies by Papalexiou and Koutsoyiannis (2012; 2016) represent perhaps the most comprehensive studies to date. However, their L-moment evaluations only evaluate the relationship between L-Skewness and L-Cv; thus they were unable to fully evaluate the goodness-of-fit of the several relatively new three-parameter pdfs introduced in their studies such as the GG and the generalized Burr type XII pdfs which would require construction of L-Kurtosis versus L-Skew diagrams which are currently unavailable for those pdfs. Analogous to those two studies, this paper uses two large scale national datasets to re-examine the question of which of the commonly used continuous distribution functions which are widely used in the fields of hydrology, meteorology and climate best fit wet-day series of observed daily precipitation data. We focus our research interest on the distribution of wet-day series of precipitation since the pdf of complete series can be derived by a mixed distribution consisting of a combination of the pdf of wet-day series and a stochastic model of the percentage and occurrence of zeros.

Instead of considering the GG distribution, the pdf recommended by both Papalexiou and Koutsoyiannis (2012, 2016), which has seen very limited use and for which analytical and/or polynomial relationships for L-Kurtosis are unavailable as they are for most commonly used pdfs in hydrology, we consider the more widely used 3 parameter generalization of the G2 distribution known as the Pearson type III (P3) distribution. Our primary objective is to use a very large national spatially distributed dataset at both the point and catchment scales, to determine a suitable probability distribution of wet-day series of daily precipitation using L-moment diagrams and probability plot correlation coefficient goodness of fit statistics.

2. Study area and data

Precipitation depths at the point and catchment scales provide important information in hydrology, meteorology, and other fields, thus our study focuses on both scales. For point precipitation, we employ a data set comprised of daily precipitation depths at 237 first-order NOAA stations from 49 U.S. states (Hawaii is excluded due to fundamentally different precipitation behavior). Station locations are shown in Figure 1a. In contrast, the areal average precipitation for 305 catchments in the international Model Parameter Estimation Experiment (MOPEX) data set (Duan et al., 2006) is also selected for analysis. The catchment locations and boundaries are shown in Figure 1b. The data were quality controlled to remove null values. When greater than 6 null values occurred in a given year or greater than 3 in a given month, the full year of data was removed. When fewer than these numbers of null values were present, they were treated as zeroes. The average record length for point precipitation depths for the 237 sites is 24,657 days (67.5 years). The distribution of record lengths corresponding to the 237 first-order NOAA stations is shown in Figure 2. The MOPEX data set consists of 56 years of areal average daily precipitation from 1948 to 2003, corresponding to a fixed record length 20,454 days for each of the 305 catchments shown in Figure 1b.

[*Figure* 1 *goes here*]

[*Figure* 2 *goes here*]

The wet-day series were extracted from both data sets. The wet-day series were constructed by excluding zero and “trace” values (those with less than 0.01 inches (approximately equivalent to 0.25 mm) recordable precipitation). Wilks (1990) discussed other ways to treat trace precipitation and left-censored data, but for convenience, they are simply excluded. The mean wet-day record lengths for point and areal average precipitation are 7,219 days (equivalent to nearly 20 years) and 14,043 days (more than 38 years), respectively. The distributions of wet-day record length are shown in Figure 3. As expected, the proportion of wet-days in the areal average precipitation data set is higher than that in the point precipitation data set.

[*Figure* 3 *goes here*]

3. Methodology

This section describes the methods of analysis used for assessing the goodness-of-fit of various distributional hypotheses, namely, L-moment diagrams and probability plot correlation coefficients.

3.1 L-Moment Diagrams

L-moment diagrams are now a widely accepted approach for evaluating the goodness of fit of alternative distributions to observations. The theory and application of L-moments introduced by Hosking (1990) is now widely available in the literature (Stedinger et al., 1993; Hosking and Wallis, 1997), hence it is not reproduced here.

The distribution of daily precipitation totals is highly skewed due to the large proportion of days with zero or small precipitation. Higher order conventional moment ratios such as skewness and kurtosis are very sensitive to extreme values and can exhibit enormous downward bias even for extremely large sample sizes (Vogel and Fennessey, 1993) as is the case here. However, L-moment ratios are approximately unbiased in comparison to conventional moment ratios, thus providing a particularly useful tool for investigating the pdf of daily wet-day precipitation series.

L-moment ratio diagrams provide a convenient graphical image to view the characteristics of sample data compared to theoretical statistical distributions. The L-moment diagrams: L-Kurtosis (τ_4) vs L-Skew (τ_3) and L-Cv (τ_2) vs L-Skew (τ_3) enable us to compare the goodness of fit of a range of three-parameter, two-parameter, and one-parameter (or special case) distributions. Table 2 displays distributions analyzed by means of the τ_4 vs τ_3 L-moment ratio diagrams.

[*Table* 2 *goes here*]

Table 3 displays distributions analyzed by means of the τ_2 vs τ_3 L-moment ratio diagrams.

[*Table* 3 *goes here*]

L-moment ratio diagrams have been used before to examine the distribution of series of annual maximum precipitation data (Pilon et al., 1991;Park and Jung, 2002;Lee and Maeng, 2003;Papalexiou and Koutsoyiannis, 2013) and left-censored records (Deidda and Puliga, 2006). Other than the two recent global studies by Papalexiou and Koutsoyiannis (2012, 2016) which examined the agreement between empirical and theoretical relationships between L-Cv and L-Skew, this is the only study we are aware of, in which a set of daily wet-day precipitation records have been subjected to such a comprehensive L-moment goodness-of-fit analysis. L-moment estimators were chosen in this study for a variety of reasons: (1) they are easily computed and nicely summarized by Hosking and Wallis (1997) for all the cases considered in this study, and (2) estimates of L-moments are unbiased and estimates of L-moment ratios are nearly unbiased, and thus for the extremely large sample sizes considered here, sampling variability of empirical L-moment ratios will be extremely small especially when contrasted to the variability among the theoretical L-moment ratios corresponding to the various distributions considered..

3.2 Probability plot correlation coefficient goodness-of-fit evaluation

Probability plots are constructed for each of the wet-day series using L-moment estimators of the distribution parameters (see Hosking and Wallis (1997)) for the distributions indicated in Table 4. A probability plot is constructed in such a manner as to ensure that the observations will appear to create a linear relationship when they arise from the hypothesized distribution assumed for each plot.

[*Table* 4 *goes here*]

The goodness of fit of each probability plot is summarized using a probability plot correlation coefficient (PPCC, or simply, r) which is simply a measure of the linearity of the plot. The PPCC statistic has a maximum value of 1. The PPCC has been shown to be a powerful statistic for evaluating the goodness-of-fit of a wide range of alternative distributional hypotheses (Stedinger et al., 1993) and for performing hypothesis tests of various two parameter distributional alternatives.

To construct a probability plot and to estimate a PPCC, requires estimation of a plotting position. There are two classes of plotting positions, those that yield unbiased exceedance probabilities and those that yield unbiased quantile estimates. The Weibull plotting position given by p=i/(n+1) yields an unbiased estimate of exceedance probability regardless of the underlying distribution (see Stedinger et al. (1993)). Alternatively there would be a unique plotting position to use for each probability distribution, and it is now well known that unbiased plotting positions for three parameter distributions require an additional parameter to estimate within the plotting position. For example, Vogel and McMartin (1991) derived an unbiased plotting position for the P3 distribution which depends upon the skewness of the distribution, a parameter which adds so much additional uncertainty to the analysis that led Vogel and McMartin (1991), after considerable analysis, to not recommend its use. To put all the distributional alternatives on the same footing, we chose to use the Weibull plotting position for estimation of all PPCC values.

4. Results and analysis

4.1 L-Moment Diagrams

4.1.1 L-Cv vs L-Skew

Figure 4 displays empirical and theoretical distributional relationships between L-Cv and L-Skew for point values of daily precipitation (Figure 4a) and areal average values of daily precipitation (Figure 4b). The various curves represent the theoretical relationship between L-Cv and L-Skew for the distributions indicated. Each plotted point represents the empirical relationship between L-Cv and L-Skew for a single precipitation station or catchment. By comparing the empirically derived points with the theoretical curves, it is possible to see the degree to which the distributional tail behavior of the data record matches those of the candidate distributions. We emphasize again, importantly, that the sample sizes are large enough in this study so that one may, approximately, ignore sampling variability in all L-moment diagrams. This phenomenon was nicely illustrated in Figure 2 of Blum et al. (2017), using synthetic data, for record lengths similar to those used here, but corresponding to daily streamflow records.

[*Figure* 4 *goes here*]

In Figure 4a, the L-moment ratios fall primarily within a region bounded by the G2 and GP2 theoretical curves, with the W2 passing through some of the points. In Figure 4b, the L-moment ratios fall primarily in the upper region of the W2 theoretical curve, with the G2 passing through or very close to most of the points. These patterns do not indicate a clearly preferred distribution for point values, especially considering that the large sample sizes associated with these series result in negligible sampling variability. However, Figure 4b documents that the G2 pdf provides a good approximation to the pdf of wet-day series for areal average values.

Blum et al. (2017, Figure 2) used L-moment diagrams for complete and synthetic series of daily streamflow observations to demonstrate that the sampling variability in L-moment ratios is negligible for the sample sizes considered in this study. Thus, the scatter shown in Figure 4 is likely due to real distributional differences rather than due to sampling variability as is often the case when one constructs L-moment diagrams for short AMS precipitation and streamflow records, as is the case in most previous studies which have employed L-moment ratio diagrams.

4.1.2 L-Kurtosis vs L-Skew

Figure 5 displays empirical and theoretical distributional relationships between L-Kurtosis vs L-Skew point values of daily precipitation (Figure 5a) and areal average values of daily precipitation (Figure 5b). The empirical relationships of plotted points for both wet-day series are very similar to the theoretical relationship for the P3 distribution. In fact, among the pdfs considered in Figure 5, the P3 pdf seems to be the only 3-parameter distribution that could possibly fit the wet-day record data. Although there is a small proportion of points lying outside the P3 curve, the overall fit is still very striking.

It should also be noted that the L-moment ratio estimates for both wet-day series occupy a space that can be well represented by the KAP distribution, which occupies a region of the L-Kurtosis vs L-Skew diagram as shown in Figure A1 of Hosking and Wallis (1997). A complete description of the 4-parameter KAP distribution is referred to Hosking (1994) and Hosking and Wallis (1997).

[*Figure* 5 *goes here*]

4.2 PPCC

4.2.1 Standard boxplots of PPCC

The L-moment ratio diagrams were useful for identifying several potential candidate distributions for representing the wet-day daily precipitation series at the point and catchment scales. From that analysis we conclude that a four parameter Kappa pdf is needed to approximate the pdf of point wet-day series whereas a G2 and P3 pdf are adequate to approximate the pdf of areal average wet-day series. The PPCC statistic offers another quantitative method for comparing the goodness of fit of different distributions to the daily precipitation observations. Table 5 summarizes the central tendency and spread of the values of PPCC for each of the distributions for the wet-day series of point and catchment scale daily precipitation, respectively. The highest values for the mean, median, 95^{th} percentile, and 5^{th} percentile of the PPCC are shown in bold type. The lowest values of the sample standard deviation of the PPCC values, denoted $\hat{s}$, are also shown in bold. Figure 6 illustrates box-plots of the values of PPCC for distributions fitted to the wet-day series of daily precipitation data at the point and catchment scales.

[*Table* 5 *goes here*]

[*Figure* 6 *goes here*]

Figure 6 and Table 5 indicate that for the wet-day-series of point daily precipitation depths, all the distributions have median PPCCs well above 0.9, but only the median PPCCs of G2, P3, and KAP distributions are over 0.99. The same situation appears in the catchment scale precipitation, except that the median PPCCs of the remaining four distributions are significantly lower than the corresponding values for point precipitation.

The insets in Figure 6 show detailed views of the boxplots of PPCC values for the G2, P3, and KAP distributions for point and areal average daily precipitation. From Figure 6a, KAP distribution results in the best goodness-of-fit for point precipitation because all of its indices are the best, while the P3 distribution generally performs better than the G2 distribution. However, for catchment-scale precipitation (Figure 6b), the four parameter KAP distribution is no longer competitive, and both the G2 and P3 pdfs will suffice. We are reluctant to advocate the use of a four parameter pdf, such as the KAP distribution, due to its inherent complexity, though such a pdf may be needed for point values as evidenced from our analyses.

4.2.2 Graphical comparison of P3, G2, and KAP

Across all previous comparisons, the P3, G2, and KAP are the best fitting distributions for describing daily precipitation at the point or catchment scales. The insets in Figure 6 identify the distributions that exhibit the best fit to the each observed series. However, these inserts do not indicate by how much the best performing distribution outperforms the second or third best. For this purpose, pairwise comparisons of the PPCC values of two highly performing distributions for all the stations and catchments are instructive. A simple graphical method can accomplish this goal.

Figure 7 compares the PPCC values of the P3 (vertical axis) and G2 (horizontal axis) distributions for point- and catchment-scale daily precipitation. Approximately 98% of stations are displayed on the figure; the remaining points lie outside the plot domains. Points lying above the diagonal line indicate that the P3 distribution has a higher PPCC for that particular station, and points lying below the diagonal line indicate the G2 results in a higher PPCC. Figure 7a shows that in nearly every case, the P3 distribution outperforms the G2 distribution. When the G2 does outperform the P3, the PPCCs are both very high and nearly equal. The point-scale precipitation plot shows that the P3 distribution performs significantly better than the G2 distribution in many cases. Thus, we conclude the P3 distribution better represents wet-day daily point precipitation than the more commonly used G2 distribution in nearly every case. Figure 7b compares the PPCC values of P3 and G2 for the catchment-scale precipitation. The results are nearly the same as for the point-scale precipitation in the sense that most points are above the diagonal line; while, for a few catchments where G2 does outperform P3, the points lie on the dividing line, showing only very slight superiority.

[*Figure* 7 *goes here*]

Figure 8 displays similar plots comparing the KAP (vertical axis) and P3 (horizontal axis) distribution for point- and catchment-scale daily precipitation. It can be seen in Figure 8a that the KAP distribution does not always outperform the P3 pdf, as one might expect given that it has an additional parameter. We are reluctant to advocate the

KAP pdf given its additional model complexity combined with the fact that it does not appear to provide a uniform improvement, in either case, over the P3 pdf.

[*Figure* 8 *goes here*]

5. Discussion

From the L-moment diagrams and PPCC comparisons we concluded that a KAP pdf is required to fully capture the tail behavior of point wet-day series, though both P3 and G2 can provide reasonable approximations in many situations. In contrast, we found that a KAP pdf is not needed to approximate the behavior of areal average wet-day series, where instead, either a P3 or G2 model would suffice. In this section we evaluate the relationship between these findings and the size of the catchments considered.

Figure 9 displays the PPCC values of P3 and G2 pdfs versus catchment drainage area for areal average wet-day series. The PPCC values are chosen from 0.99-1, approximately 96% of catchments are displayed on the figure; the remaining points lie outside the plot domains. It can be seen that for most of the catchments, the PPCC values for G2 and P3 pdfs are very close, with points corresponding to G2 and P3 pdfs almost overlapping. This is especially true for PPCC values higher than 0.998. The phenomena clearly indicates that when G2 can well represent the behavior of catchment-scale wet-day precipitation series, P3 also provides very good performance. However, for the areas where PPCC values are lower than 0.996, the P3 distribution outperforms the G2 distribution for most cases, with a very slight improvement.

[*Figure* 9 *goes here*]

Figure 10 shows the spatial map of catchments with the corresponding best distribution functions for areal average wet-day series. KAP distribution is the best pdf for large proportion of the catchments especially in the middle of US. P3 distribution occupies the second large proportion of the catchments especially in east-central US. Only a very few catchments can be best represented by G2 distribution. Seen from Figure 10, it seems that the performances of the three pdfs vary greatly. However, as we have seen from previous figures, the differences between the three pdfs for catchments are very small.

[*Figure* 10 *goes here*]

6. Conclusions

This study has demonstrated that L-moment diagrams and probability plot correlation coefficient goodness of fit evaluations can provide new insight into the distribution of very long series of daily wet-day precipitation at both the point and catchment scales. Although previous studies have claimed that the commonly used 2-parameter Gamma distribution performs fairly well on the basis of traditional goodness-of-fit tests, this study reveals, through the use of L-moment diagrams and probability plot correlation coefficient goodness of fit evaluations that very long series of uncensored daily point and areal average precipitation are better approximated by a KAP distribution and a Pearson-III distribution respectively, and importantly, they do not resemble any of the other commonly used distributions. Analogous to the recent study by Papalexiou and Koutsoyiannis (2016), our evaluations yield very different conclusions than previous research on this subject and thus could have important implications in climate change investigations and other studies which employ a pdf of daily precipitation.

We conclude that for representing wet-day precipitation, the Gamma and Pearson-III distributions are comparable with the 4-parameter Kappa distribution for the areal average precipitation; however, when the point precipitation is of concern, the Kappa distribution should be the distribution of choice. We also conclude that future investigations should consider comparisons between the generalized Gamma distribution introduced by Papalexiou and Koutsoyiannis (2012, 2016) for wet-day daily precipitation and the G2, Pearson type III and Kappa distributions recommended here.

Once analytical and polynomial L-moment relationships and parameter estimation methods become available for the GG distribution, future studies should compare the P3 and GG distributions on wet-day series, because on the basis of this study, and Papalexiou and Koutsoyiannis (2016), the P3 and GG distributions appear to have tremendous potential for approximating the distribution of wet-day series.

Acknowledgements

The first and third authors are partially supported by the National Natural Science Foundation of China (No. 91547116, 51709033, 91647201). Special thanks are given to Dr. Simon M. Papalexiou and other two anonymous reviewers and editors for their constructive remarks.

References

Blum, A. G., Archfield, S. A., and Vogel, R. M.: On the probability distribution of daily streamflow in the United States, Hydrology and Earth System Sciences, 21, 3093, 2017.

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D.: Precipitation-frequency atlas of the United States, NOAA atlas, 14, 2006.

Buishand, T. A.: Some remarks on the use of daily rainfall models, Journal of Hydrology, 36, 295-308, 1978.

Burgueno, A., Martinez, M. D., Lana, X., and Serra, C.: Statistical distributions of the daily rainfall regime in Catalonia (northeastern Spain) for the years 1950 2000, International Journal of Climatology, 25, 1381-1403, 2005.

Chen, J., and Brissette, F. P.: Stochastic generation of daily precipitation amounts: review and evaluation of different models, Climate Research, 59, 189-206, 2014.

Deidda, R., and Puliga, M.: Sensitivity of goodness-of-fit statistics to rainfall data rounding off, Physics and Chemistry of the Earth, Parts A/B/C, 31, 1240-1251, 2006.

Duan, J., Sikka, A. K., and Grant, G. E.: A comparison of stochastic models for generating daily precipitation at the HJ Andrews Experimental Forest, 1995.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., and Hay, L.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3-17, 2006.

Easterling, D. R., Evans, J., Groisman, P. Y., Karl, T. R., Kunkel, K. E., and Ambenje, P.: Observed variability and trends in extreme climate events: a brief review, Bulletin of the American Meteorological Society, 81, 417-425, 2000.

Geng, S., de Vries, F. W. P., and Supit, I.: A simple method for generating daily rainfall data, Agricultural and Forest Meteorology, 36, 363-376, 1986.

Groisman, P. Y., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C. M., Wibig, J., and Fortuniak, K.: Changes in the probability of heavy precipitation: important indicators of climatic change, in: Weather and Climate Extremes, Springer, 243-283, 1999.

Hershfield, D. M.: Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, 1961.

Hosking, J. R.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, Journal of the Royal Statistical Society. Series B (Methodological), 105-124, 1990.

Hosking, J. R.: The four-parameter kappa distribution, IBM Journal of Research and Development, 38, 251-258, 1994.

Hosking, J. R. M., and Wallis, J. R.: Regional frequency analysis : an approach based on L-moments, Cambridge University Press, 1997.

Karl, T. R., and Knight, R. W.: Secular trends of precipitation amount, frequency, and intensity in the United States, Bulletin of the American Meteorological society, 79, 231-241, 1998.

Kigobe, M., McIntyre, N., Wheater, H., and Chandler, R.. Multi-site stochastic modelling of daily rainfall in Uganda. Hydrological sciences journal, 56, 17-33, 2011.

Lee, S. H., and Maeng, S. J.: Frequency analysis of extreme rainfall using L moment, Irrigation and Drainage, 52, 219-230, 2003.

Li, Z., Brissette, F., Chen, J.. Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds. Hydrological Processes, 27: 3718-3729, 2013.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T.: Review of trend analysis and climate change projections of extreme precipitation and floods in Europe, Journal of Hydrology, 519, 3634-3650, 2014.

Mehrotra, R., Srikanthan, R., and Sharma, A.: A comparison of three stochastic multi-site precipitation occurrence generators, Journal of Hydrology, 331, 280-292, 2006.

Naghavi, B., and Yu, F. X.: Regional frequency analysis of extreme precipitation in Louisiana, Journal of Hydraulic Engineering, 121, 819-827, 1995.

Papalexiou, S. M., and Koutsoyiannis, D.: Entropy based derivation of probability distributions: A case study to daily rainfall, Advances in Water Resources, 45, 51-57, 2012.

Papalexiou, S. M., and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resources Research, 49, 187-201, 2013.

Papalexiou, S. M., and Koutsoyiannis, D.: A global survey on the seasonal variation of the marginal distribution of daily precipitation, Advances in Water Resources, 94, 131-145, 2016.

Park, J.-S., and Jung, H.-S.: Modelling Korean extreme rainfall using a Kappa distribution and maximum likelihood estimate, Theoretical and Applied climatology, 72, 55-64, 2002.

Pilon, P. J., Adamowski, K., and Alila, Y.: Regional analysis of annual maxima precipitation using L-moments, Atmospheric Research, 27, 81-92, 1991.

Schoof, J. T., Pryor, S. C., and Surprenant, J.: Development of daily precipitation projections for the United States based on probabilistic downscaling, Journal of Geophysical Research: Atmospheres, 115, D13, 2010.

Shoji, T., and Kitaura, H.: Statistical and geostatistical analysis of rainfall in central Japan, Computers & Geosciences, 32, 1007-1024, 2006.

Srikanthan, R., and McMahon, T.: Stochastic generation of annual, monthly and daily climate data: A review, Hydrology and Earth System Sciences Discussions, 5, 653-670, 2001.

Stedinger, J. R., R.M. Vogel and E. Foufoula-Georgiou: Frequency analysis of extreme events, Handbook of Hydrology, Chapter 18, McGraw Hill Book Co, D.R. Maidment - editor in chief, 1993..

Thom, H. C.: A frequency distribution for precipitation, Bulletin of the American Meteorological Society, 32, 397, 1951.

Trenberth, K. E.: Changes in precipitation with climate change, Climate Research, 47, 123-138, 2011.

Vogel, R. M., and Fennessey, N. M.: L moment diagrams should replace product moment diagrams, Water Resources Research, 29, 1745-1752, 1993.

Vogel, R. W., and McMartin, D. E.: Probability Plot Goodness‐of‐Fit and Skewness Estimation Procedures for the Pearson Type 3 Distribution, Water resources research, 27, 3149-3158, 1991.

Waggoner, P.E., 1989: Anticipating the frequency distribution of precipitation if climate change alters its mean, Agricultural and Forest Meteorology, 47, 321 – 337.

Watterson, I., and Dix, M.: Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution, Journal of Geophysical Research: Atmospheres, 108, 2003.

Watterson, I. G.: Simulated changes due to global warming in the variability of precipitation, and their interpretation using a gamma-distributed stochastic model, Advances in Water Resources, 28, 1368-1381, 2005.

Waymire, E., and Gupta, V. K.: The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models, Water resources research, 17, 1261-1272, 1981.

Wilby, R. L., and Wigley, T.: Future changes in the distribution of daily precipitation totals across North America, Geophysical Research Letters, 29, 2002.

Wilks, D. S.: Maximum likelihood estimation for the gamma distribution using data containing zeros, Journal of Climate, 3, 1495-1501, 1990.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, Journal of Hydrology, 210, 178-191, 1998.

Wilks, D. S., and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Progress in physical geography, 23, 329-357, 1999.

Wilson, P. S., and Toumi, R.. A fundamental probability distribution for heavy rainfall. Geophysical Research Letters, 32, L14812, 2005.

Woolhiser, D. A., and Roldan, J.: Stochastic daily precipitation models: 2. A comparison of distributions of amounts, Water resources research, 18, 1461-1468, 1982.

Yoo, C., Jung, K. S., and Kim, T. W.: Rainfall frequency analysis using a mixed Gamma distribution: evaluation of the global warming effect on daily rainfall, Hydrological Processes, 19, 3851-3861, 2005.

Table captions:

Table 1: Review of literature pertinent to daily precipitation probability distribution selection.

Table 2: Table 2: Theoretical probability distributions presented on the L-Kurtosis vs L-Skew L-moment diagram. *Italicized distributions are special cases of other distributions*.

Table 3: Theoretical probability distributions presented on the L-Cv vs L-Skew L-moment diagram.

Table 4: Distributions used in probability plot goodness of fit evaluations.
Table 5: Central tendency and spread of values of PPCC for the 237 precipitation stations 305 areal average precipitation catchments.

Figure captions:

Figure 1: Map showing locations of a) 237 precipitation gaging stations, and b) 305 catchments.

Figure 2: Distribution of full record length of point precipitation base on weather stations.

Figure 3: Distribution of wet-day record length: a) point precipitation; and b) areal average precipitation over watersheds. Days with zero precipitation are removed in the wet-day records

Figure 4: L-Cv vs L-Skew L-moment ratio diagram of sample L-moments and theoretical distributions: a) point precipitation; and b) areal average precipitation depths.

Figure 5: L-Skew vs L-Kurtosis L-moment ratio diagram of sample L-moments and theoretical distributions: a) point precipitation; and b) areal average precipitation depths. Logistic (L), Normal (N), Uniform (U), Gumbel (G), and Exponential (E) distributions appear as a single point.

Figure 6: Standard boxplots of r for all 7 distributions evaluated for wet-day series of a) point precipitation, and b) areal average precipitation depths.

Figure 7: Comparison of PPCC (r) values for the P3 (vertical axis) and G2 (horizontal axis) distributions for the a) point, and b) areal average precipitation depths series. Points lying above the line represent stations with a higher r for the P3 distribution than G2 distribution.

Figure 8: Comparison of r values for P3 (horizontal axis) and KAP (vertical axis) distributions for the a) point and b) areal average precipitation depths' wet-day series.

Figure 9: The spatial distribution of best daily precipitation distribution function.

Figure 10: The spatial map of catchments with the corresponding best distribution functions for areal average wet-day series.

Tables

Table 1: Review of literature pertinent to daily precipitation probability distribution selection.

1. Stochastic Precipitation Modelling:

Author	Year	Stations	Series type	Duration	Distribution	Justification
Thom	1951		Wet-day	1-day	Gamma	
Buishand	1978	6	Wet-day	1-day	Gamma	Cv-Cs ratio
Geng et al	1986	6	Wet-day, by month	1-day, monthly	Gamma	Regress. fit: β vs mean wet-day depth
Woolhiser and Roldan	1982		Wet-day	1-day	Mixed Exponential	MLE, Akaike Information Criterion
Duan et al	1995	1	Wet-day, by month	1-day	Calib. W2, Gamma	MLE, Chi-sq test
Wilks	1998	25	Wet-day	1-day	Mixed Exponential	MLE, goodness of fit
Waterson and Dix	2003		Wet-day	1-day	Gamma	Literature
Burgueno et al	2005	75	Wet-day	1-day	Exponential, Weibull	Normalized Rainfall Curve
Kigobe et al	2011	110	Wet-day, by month		Gamma	
Li et al	2013	24	Wet-day	1-day	Mixed Exponential	Goodness of fit and Kolmogorov–Smirnov tests
Schoof	2015	Grided precipitation	Wet-day	1-day	Gamma	Goodness of fit

2. Precipitation Frequency Analysis

Author	Year	Stations	Series type	Duration	Distribution	Justification
Hershfield (TP-40)	1962		AMS	24 hour	Gumbel	
Pilon et al	1991	75	AMS	5 min - 24 hour	GEV	L-moments
Naghavi & Yu	1995	25	AMS	1-24 hour	GEV	L-moments, PWMs, Monte Carlo experiments
Park and Jung	2002	61	AMS	1, 2-day	Kappa(4)	
Lee and Maeng	2003	38	AMS	1-day	GEV, GLO	L-moments
Bonnin et al	2006		AMS	5 min - 24 hour	GEV	L-moments
Shoji and Kitaura	2006	243	Complete, Wet-day	Hour, Day, Month, Year	Lognormal, Weibull	Goodness of fit
Deidda and Puliga	2006	200	Left Censored Wet-day PDS	1-day	Generalized Pareto	“Failure-to-reject” method, L-moments
Wilson and Toumi	2005	270	Complete	1-day	Self-derived	
Papalexiou and Koutsoyiannis	2012	11,519	Wet-Day	1-day	Generalized Gamma	L-moments
Papalexiou and Koutsoyiannis	2013	15,137	AMS	1-day	GEV	L-moments
Papalexiou and Koutsoyiannis,	2016	14,157	Wet-Day, by month	1-day	Generalized Gamma and Burr type XII	L-moments and Goodness-of-fit

3. Precipitation Trends and Climate Change						
Author	**Year**	**Stations**	**Series type**	**Duration**	**Distribution**	**Justification**
Waggoner	1989	55	Monthly	1-month	Gamma	Literature Review
Groisman et al	1999	1313	Summer (wet-day)	1-day	Gamma	Literature Review, goodness of fit to extreme rainfall quantiles
Wilby and Wigley	2002	GCM	Seasonal	1-day	Gamma	Literature Review
Yoo et al	2005	31	Monthly (wet-day)	1-day	Gamma	Literature Review
Watterson	2005	GCM	January, July	1-month (daily forced)	Gamma	Literature Review

Table 2: Theoretical probability distributions presented on the L-Kurtosis vs L-Skew L-moment diagram. *Italicized distributions are special cases of other distributions.*

Distribution	Abbreviation	PDF	Parameters
Generalized Extreme Value Type III	GEV	$f(x)=\frac{1}{\eta}\left[1-\left(\frac{x-\omega}{\eta}\right)\beta\right]^{\frac{1}{\beta}-1}\exp^{\left\{-\left[1-\left(\frac{x-\omega}{\eta}\right)\beta\right]^{\frac{1}{\beta}}\right\}}$	3
Generalized Logistic	GLO	$f(x)=\frac{be^{-\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{-\frac{x-\mu}{\sigma}}\right)^{b+1}}$	3
Generalized Pareto	GPA	$f(x)=\frac{1}{\sigma}\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{(-1/\xi-1)}$	3
Lognormal	LN3	$f(x)=\frac{1}{\sqrt{2\pi}(x-\alpha)\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(\ln(x-\alpha)-\mu^2\right)\right]$	3
Pearson Type III	P3	$f(x)=\frac{p/a^d}{\Gamma(d/p)}x^{d-1}e^{-(x/a)^p}$	3
Exponential	E	$f(x)=\begin{cases}\lambda e^{-\lambda x}, x\geq 0\\ 0, x<0\end{cases}$	2
Gumbel	G	$f(x)=\frac{1}{\beta}e^{-\left(z+e^{-z}\right)}, \mathrm{Z}=\frac{x-\mu}{\beta}$	2
Normal	N	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	2
Logistic	L	$f(x)=\frac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)^2}$	2
Uniform	U	$f(x)=\begin{cases}\frac{1}{b-a}, a<x<b\\ 0, x<a \text{ or } x>b\end{cases}$	1

Table 3: Theoretical probability distributions presented on the L-Cv vs L-Skew L-moment diagram.

Distribution	Abbreviation	PDF	Parameters
Gamma	G2	$f(x)=\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$	2
Generalized Pareto	GP2	$f(x)=\frac{1}{\sigma}\left(1+\frac{\xi x}{\sigma}\right)^{(-1/\xi-1)}$	2
Lognormal	LN2	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	2

Weibull	W2	$f(x)=\frac{k}{\phi}\left(\frac{x}{\phi}\right)^{k-1}\exp^{\left\{-\left[\frac{k}{\phi}\right]^{k}\right\}}$	2

Table 4: Distributions used in probability plot goodness of fit evaluations.

Distribution	Abbreviation	Parameters
Generalized Extreme Value Type III	GEV	3
Generalized Logistic	GLO	3
Generalized Pareto	GPA	3
Lognormal	LN3	3
Pearson Type III	P3	3
Gamma	G2	2
Kappa	KAP	4

Table 5: Central tendency and spread of values of PPCC for the 237 precipitation stations and 305 catchments.

Distribution	Point precipitation			*Percentiles*		Areal average precipitation			*Percentiles*	
	Mean	Median	$\hat{s}$	95th	5th	Mean	Median	$\hat{s}$	95th	5th
P3	0.9952	0.9971	0.0063	0.9995	0.9872	**0.9977**	0.9985	**0.0028**	0.9996	**0.9936**
GEV	0.9338	0.9375	0.0222	0.9609	0.8944	0.8003	0.7965	0.0474	0.8917	0.7264
GPA	0.9793	0.9828	0.0145	0.9949	0.9500	0.8688	0.8687	0.0484	0.9586	0.7894
GLO	0.9115	0.9154	0.0235	0.9423	0.8734	0.7800	0.7750	0.0441	0.8669	0.7101
LN3	0.9838	0.9855	0.0075	0.9924	0.9727	0.9362	0.9373	0.0224	0.9737	0.8983
G2	0.9925	0.9949	0.0079	0.9990	0.9789	0.9974	0.9985	0.0034	0.9996	0.9924
KAP	**0.9971**	**0.9985**	**0.0048**	**0.9997**	**0.9915**	0.9976	**0.9987**	0.0026	**0.9998**	0.9929

Figures

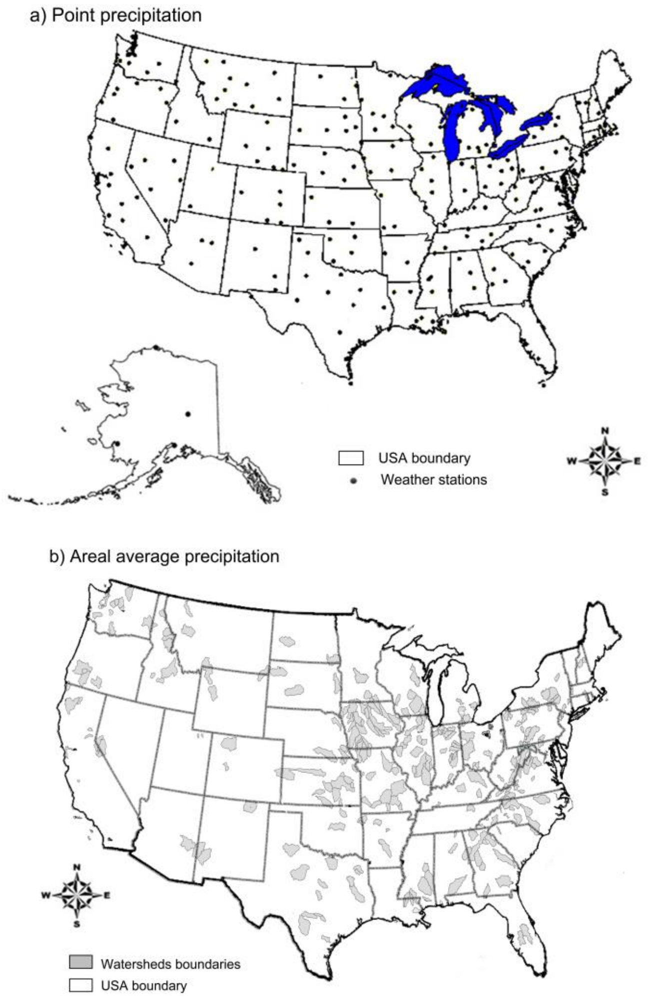

Figure 1: Map showing locations of a) 237 point precipitation gaging stations, and b) 305 MOPEX catchments.

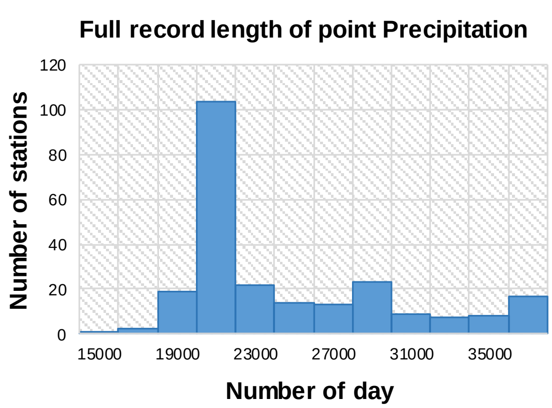

Figure 2: Distribution of length of records of point daily precipitation data for the 237 gaging stations depicted in Figure 1a.

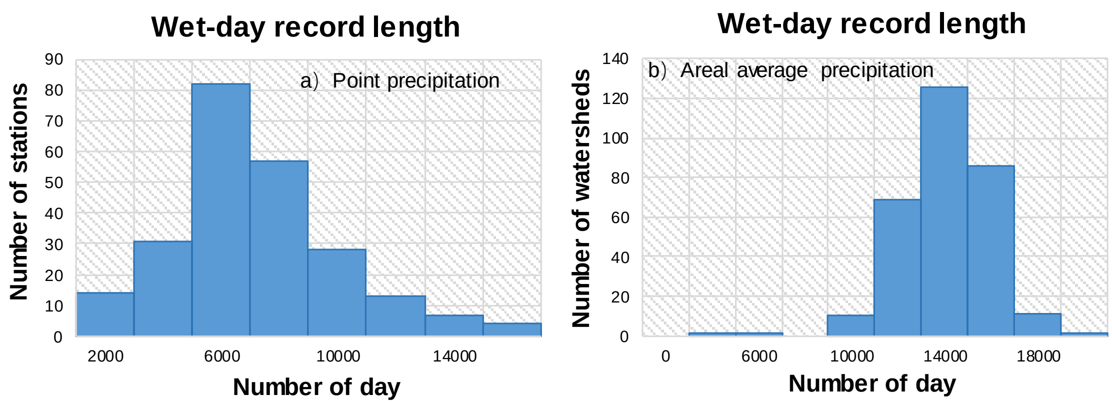

Figure 3: Distribution of wet-day record lengths corresponding to the two datasets: a) point precipitation; and b) areal average precipitation over catchments. Days with zero precipitation are removed in the wet-day records

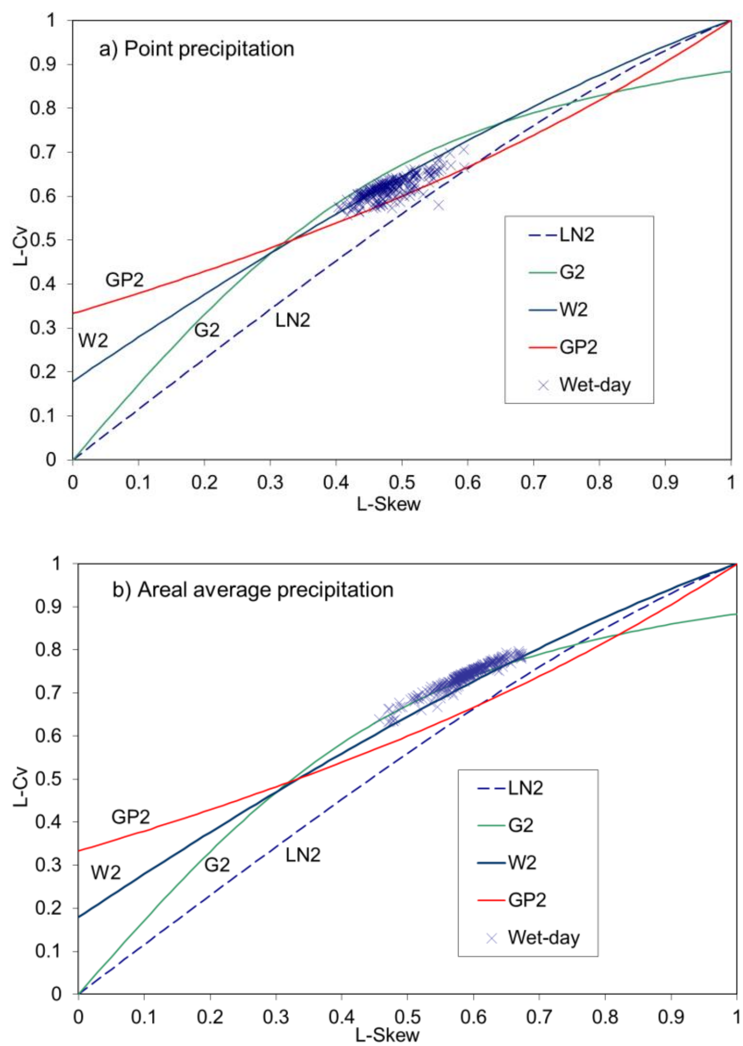

Figure 4: L-Cv vs L-Skew L-moment ratio diagram of sample L-moments and theoretical distributions for: a) point daily precipitation; and b) areal average daily precipitation depths.

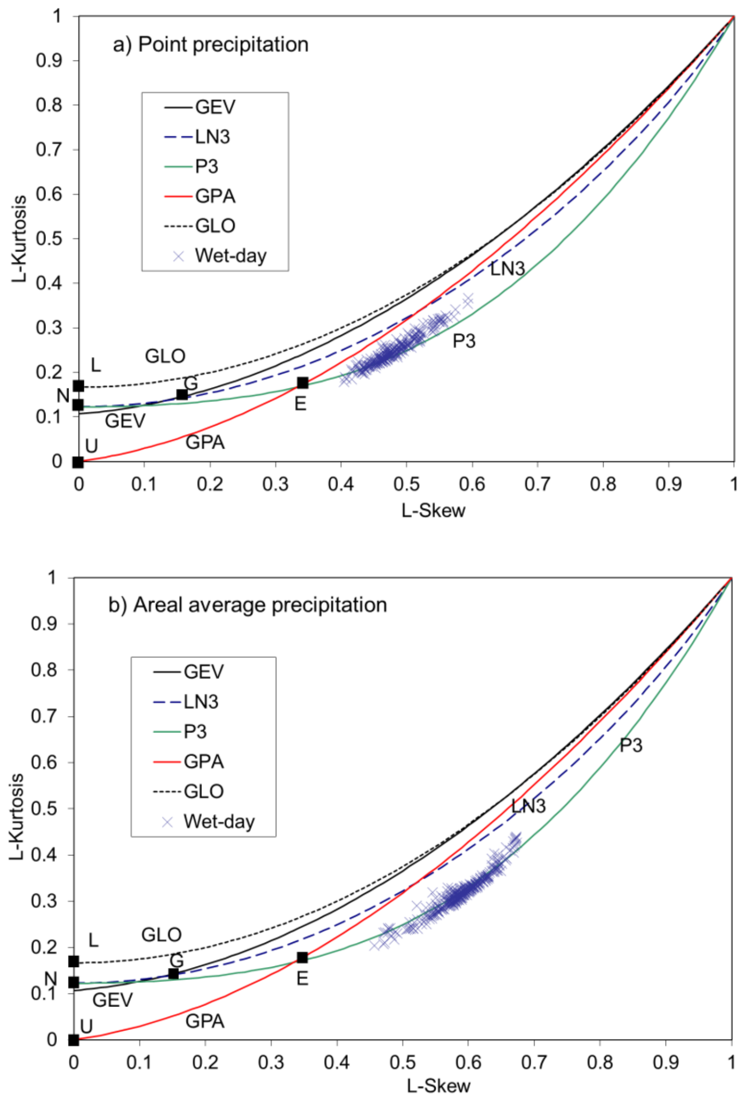

Figure 5: L-Skew vs L-Kurtosis L-moment ratio diagram of sample L-moments and theoretical distributions for: a) point daily precipitation; and b) areal average daily precipitation depths. Note that Logistic (L), Normal (N), Uniform (U), Gumbel (G), and Exponential (E) distributions appear as a single point.

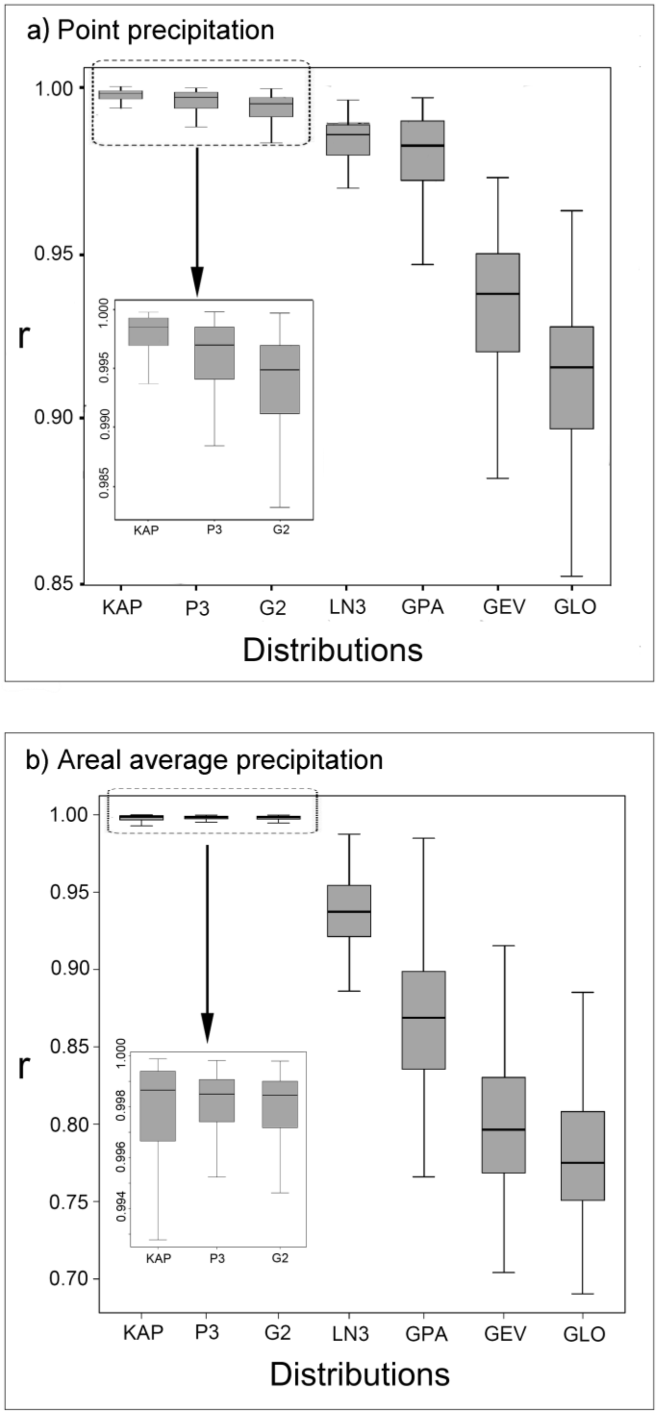

Figure 6: Standard boxplots of r for all 7 distributions evaluated for wet-day series of a) point precipitation, and b) areal average precipitation depths.

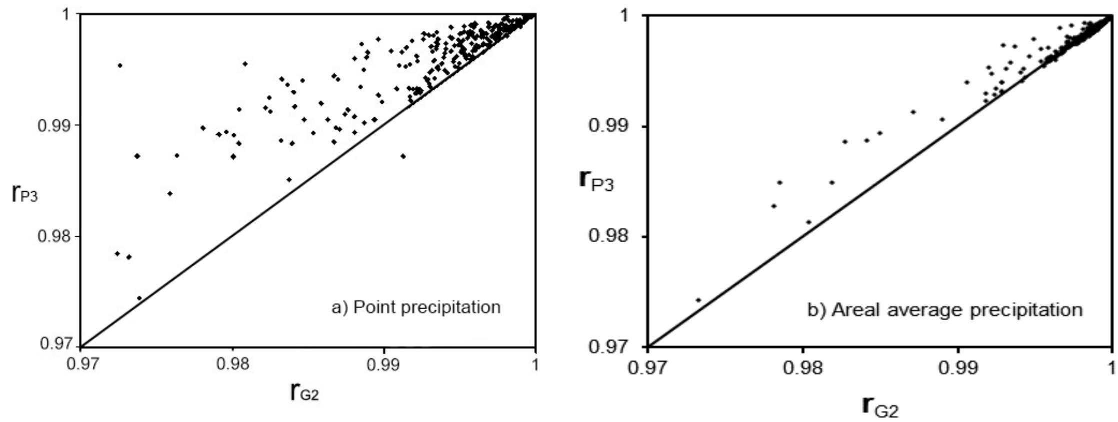

Figure 7: Comparison of PPCC (r) values for the P3 (vertical axis) and G2 (horizontal axis) distributions for the a) point, and b) areal average precipitation depths series. Points lying above the line represent stations with a higher r for the P3 distribution than G2 distribution.

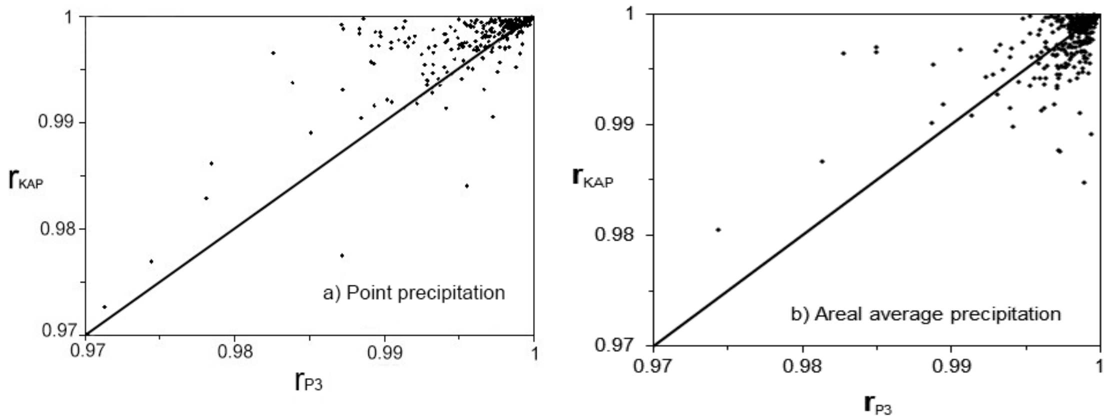

Figure 8: Comparison of r values for P3 (horizontal axis) and KAP (vertical axis) distributions for the a) point and b) areal average precipitation depths' wet-day series.

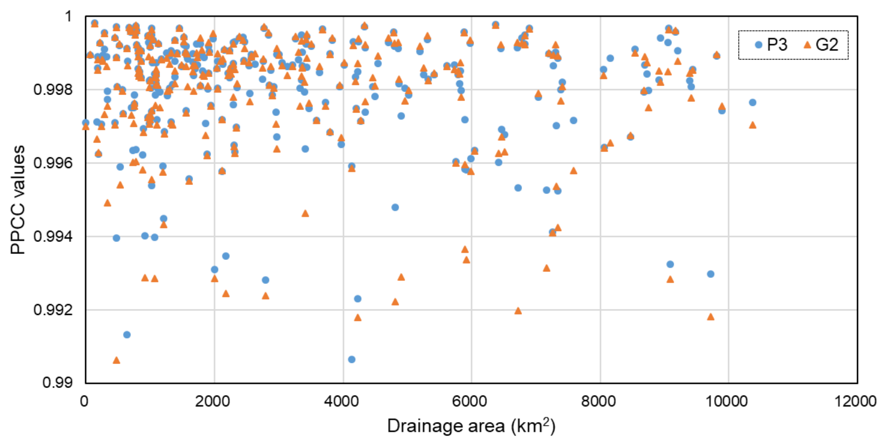

Figure 9: the PPCC values of P3 and G2 pdfs versus catchment drainage area for areal average wet-day series.

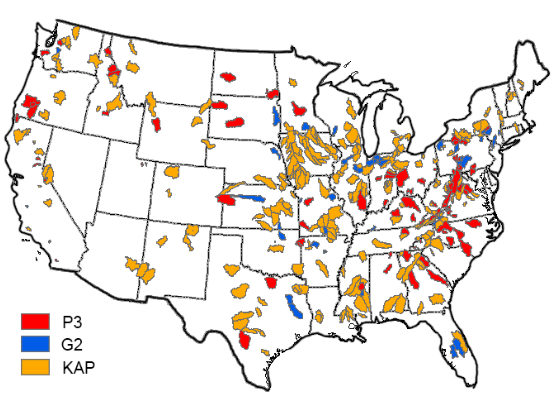

Figure 10: The spatial map of catchments with the corresponding best distribution functions for areal average wet-day series.