In search of the probability distribution of daily streamflow

Annalise G. Blum [1,2], Stacey A. Archfield [2], Richard M. Vogel[1]

[1]Civil and Environmental Engineering, Tufts University, Medford, MA, 02155, USA

[2]U.S. Geological Survey, Reston, VA, 20192, USA

Correspondence to: Stacey Archfield (sarch@usgs.gov)

Abstract. Daily streamflows are often represented by flow duration curves (FDCs), which illustrate the frequency with which flows are equaled or exceeded. FDCs have had broad applications across both operational and research hydrology for decades; however, modeling FDCs has proven elusive. Daily streamflow is a complex time series with flow values ranging over many orders of magnitude. Nevertheless, the identification of a probability distribution that can approximate daily streamflows should enhance our understanding of the behavior of daily streamflow and our ability to estimate FDCs at ungaged river locations. Comparisons of modeled and empirical FDCs at nearly 400 unregulated, perennial streams illustrates that the four-parameter kappa distribution provides a very good representation of daily streamflow across the majority of physiographic regions in the conterminous United States (US). Further, for some regions of the US, the three-parameter generalized Pareto and lognormal distributions also provide a good approximation to FDCs. Similar results are found for both period of record FDCs, representing the long-term hydrologic regime at a site, and median annual FDCs, representing the behavior of flows in a typical year.

1 Introduction

Daily streamflows are often represented by flow duration curves (FDCs), which illustrate the frequency with which flows are equaled or exceeded. FDCs have important applications including water allocation, wastewater management, hydropower assessments, sediment transport, protection of ecosystem health, and the generation of time series of daily streamflows (Archfield and Vogel, 2010; Castellarin et al., 2013; Smatkin, 2001; Vogel and Fennessey, 1995). Broad regions of the world have insufficient records of streamflow and, despite a decade of work focused on such ungaged and partially gaged basins, accurate prediction of streamflow in these locations remains a challenge (Sivapalan et al., 2003; Hrachowitz et al., 2013). Identification of a probability distribution of daily streamflows would be instrumental to the prediction of flows in ungaged basins. The goal of this study is to assess whether a single probability distribution can adequately approximate, across very broad hydroclimatic regimes, the distribution of daily streamflows, as represented by a period-of-record FDC (FDC_{POR}), which reflects the long-term or steady-state hydrologic regime at a site. Our assessment is performed at the sub-continental scale to enable consideration of a broad range of hydrologic conditions that may be experienced in practice.

Methods to predict the FDC_{POR} in ungaged basins generally fall into one of two categories: process-based or statistical. For an extensive review of these methods, refer to Chapter 7 in the book "Runoff prediction in ungaged basins" (Castellarin et al., 2013). Process-based models are an increasingly popular method of estimating FDCs at ungaged basins because they offer the ability to relate physical watershed characteristics to streamflow regimes. While promising for regions without any streamflow data, process-based FDC_{POR} models require numerous assumptions regarding runoff and climate mechanisms (Basso et al., 2015, Botter et al., 2008; Doulatyari et al., 2005, Müller and Thompson, 2016; Schaefli et al., 2013; Yokoo and Sivapalan, 2011).

Historically, most studies predicting FDC_{POR} at ungaged sites have used statistical methods, such as regression and index-flow methods, due to their parsimony and relative ease of use in operational hydrology (Castellarin et al., 2013). Yet, daily streamflow observations exhibit a very high degree of serial correlation, seasonality and other complexities and are thus neither independent nor identically distributed. Klemeš (2000) warned that ignoring these complexities can be problematic, particularly if the FDC_{POR} is used to extrapolate upper tails of the distribution. Furthermore, the fact that daily streamflows often range over many orders of magnitude presents a considerable challenge to the identification of an appropriate distribution. While multiple parameters are needed to describe the complex distribution of daily streamflows, it is also important that the model be parsimonious, because each additional parameter can hinder estimation, parameter identifiability and interpretation (Castellarin et al., 2007).

Despite these theoretical and practical challenges, there is a relatively large literature which has sought to approximate the distribution of daily streamflow with a single probability distribution for very practical purposes. The main motivations have been estimation of FDCs at ungaged sites, often based on an index-flow method (Castellarin et al., 2004, 2007; Fennessey and Vogel, 1990; Li et al., 2010; Mendicino and Senatore, 2013; Rianna, 2011; Viola et al., 2011) or for estimation of time series of daily streamflow at ungaged sites (Fennessey, 1994; Smatkin and Masse, 2000; Archfield and Vogel, 2010). A number of distributions have been proposed to describe daily streamflow. Li et al (2010) found that the three-parameter lognormal distribution (LN) adequately represented the FDC_{POR} for the southeastern Australia region. In Italy, both the four-parameter kappa (KAP) and the generalized Pareto (GPA), a special case of KAP, have been used to described

FDC_{POR} in index-flow studies (Castellarin et al., 2004, 2007; Mendicino and Senatore, 2013). Similarly, both GPA and KAP provided a good approximation for FDC_{POR} in the northeastern United States (US) (Archfield, 2009; Fennessey, 1994; Vogel and Fennessey, 1993). However, Archfield (2009) highlighted challenges in fitting both KAP and GPA to tails of the FDC_{POR}, noting these fitted distributions often exhibit lower bounds that can result in the generation of negative flows. Multiple authors have noted that a complex distribution with at least four parameters is needed to approximate the probability distribution of daily streamflows (Archfield, 2009; Castellarin et al., 2004; LeBoutillier and Waylen, 1993).

Given the challenge of selecting a single distribution to approximate the probabilistic behavior of daily streamflows, some studies have focused on only a portion of the FDC_{POR}, such as flows below the median (Fennessey and Vogel, 1990) or above the mean (Segura et al., 2013). Others have studied the distribution of streamflows by season. For eight rivers across the US, Bowers et al. (2012) developed a method to identify wet and dry season FDCs and found discharge data in wet seasons to be well-approximated by a lognormal distribution, but dry season flows sometimes better fit with a power law distribution. That study also illustrated the challenges of conducting comprehensive seasonal analyses; the authors found differences between the behavior of rivers depending on how seasons were defined as well as which distributions best fit flows in each season, suggesting that seasonal analysis of this kind may be highly site-specific. A couple of papers have documented attempts to fit a probability distribution to a mean annual FDC or a median annual FDC (FDC_{MED}), both hypothetical FDCs that express the likelihood of daily streamflow being exceeded during a typical year (Fennessey, 1994; LeBoutillier and Wayland, 1993). The FDC_{MED}, introduced by Vogel and Fennessey (1994), has a number of applications from ecology to hydropower (Lang et al., 2004; Müller et al., 2014; Kroll et al. 2015). FDC_{MED}, are increasingly common and enable the computation of tolerance or uncertainty intervals along with associated hypothesis tests for flow alteration (see Kroll et al. 2015).

To address the practical goal of estimating FDCs, this study aims to determine whether or not an existing probability distribution is capable of approximating the distribution of daily streamflows for nearly 400 perennial rivers with near-natural streamflow across the conterminous US. Differences in the performance of the probability distributions in approximating FDC_{POR} are compared across physiographic regions of the US to illustrate where these methods might be most successful. In addition, this study also considers the ability of a single probability distribution to represent the FDC_{MED}.

The paper is organized as follows. First, the method to construct an FDC_{POR} is described and the goodness-of-fit (GOF) metrics and study region are introduced. The results are then presented including L-moment ratio diagrams, and quantitative GOF comparisons among the fitted probability distributions. These GOF results are then compared by physiographic region within the US and the FDC_{MED} results are shown. Finally, the conclusion summarizes study findings and provides directions for future research.

2 Methods

2.1 *FDC estimation*

An empirical FDC_{POR} is constructed by ranking daily streamflows from all recorded years and plotting them against an estimate of their exceedance probability, known as a plotting position (Vogel and Fennessey, 1994). An FDC is defined as the complement of the cumulative distribution function:

$$1 - F_Q(q), \text{ where } F_Q(q) = P\{Q \leq q\} \tag{1}$$

where q represents observed streamflow and $F_Q(q)$ is the empirical cumulative distribution function of observed streamflow. The first step in constructing an FDC_{POR} is to rank the flows, q_i, in ascending order as in $q_{(1)} \ldots q_{(365n)}$ where n is the number of years of record. For leap years, flows from February 29 were removed to maintain consistent sample sizes across years. To obtain the probability with which each flow is exceeded, the Weibull plotting position was used, as it provides an unbiased estimate of exceedance probability, regardless of the underlying probability distribution of the ranked observations (Vogel and Fennessey, 1994):

$$P\{Q > q\} = 1 - \frac{i}{365n+1} \tag{2}$$

where i represents the rank. Vogel and Fennessey (1994) review several alternative nonparameteric plotting positions for constructing empirical FDCs at a gaged site, some of which are preferred for smaller samples. The Weibull plotting position is selected here given the large sample sizes considered (at least 40 years of daily data leading to sample sizes greater than 40x365=14600).

2.2 *Selection of candidate distributions*

As an initial assessment, L-moment ratio diagrams were used to narrow the pool of potential candidate probability distributions. L-moments are linear combinations of probability-weighted moments (Hosking and Wallis 1997), and estimates of L-moment ratios exhibit substantially less bias than moment ratio estimators, and are resistant to the influence of data outliers (Hosking and Wallis, 1997). The advantages of using L-moment diagrams in distribution identification are described in Vogel and Fennessey (1993) and Hosking and

Wallis (1997). L-moments can be directly related to ordinary product moments of a probability distribution.

Theoretical relationships between L-moment ratios have been determined for a wide class of probability distributions (Hosking and Wallis, 1997). These relations can be plotted on an L-moment ratio diagram with L-moment ratios estimated from the daily streamflows to provide a visual method of comparing various probability distributions to observed data. Vogel and Fennessey (1993) demonstrate that L-moment ratio diagrams are often superior to ordinary moment ratio diagrams, even for extremely long records of highly skewed samples of daily streamflows, as is the focus of this study. Even when parent distributions are complex, L-moment ratio diagrams are useful in identifying simpler distributions that fit the observed data sufficiently well (Stedinger et al., 1993). For a description of the theory of L-moments, see Hosking (1990).

2.3 *Goodness of fit evaluation*

To evaluate the suitability of a model to reproduce observations, a measure of the standardized mean square error commonly referred to as Nash-Sutcliffe Efficiency (NSE) is used. The most common estimator of NSE at each site is:

$$NSE = 1 - \frac{\sum_{x=1}^{X}(Q_x - Q_x^{pred})^2}{\sum_{x=1}^{X}(Q_x - \overline{Q_x})^2} \tag{3}$$

where Q_x represents observed flow at quantile x, Q_x^{pred} predicted flow at quantile x, $\overline{Q_x} = \sum_{x=1}^{X}\frac{Q_x}{X}$ the mean value of the observed flows, and X the total number of daily flows (and therefore number of quantiles). NSE values range from -∞ to a maximum of 1, which here would indicate that the estimated flows matched observed flows exactly. Because NSE are heavily influenced by the highest flows, the natural logarithms of the flows are used in the computation of NSE and herein referred to as LNSE.

Visual comparisons of the estimated and observed FDC_{POR} for candidate distributions are also presented to relate LNSE values to visual depictions of the GOF. Part of the reason why FDC_{POR} are so widely used in practice is that they provide a graphical illustration of the complete relationship between the magnitude and frequency of streamflow. Lastly, error duration curves (as in Müller and Thompson, 2016) are given for each candidate distribution to illustrate how the error is distributed across exceedance probabilities. Error is measured by calculating the ratio of observed to predicted flows across all sites.

3 *Study region and streamgages*

Only gages considered to represent near-natural streamflow conditions (as identified by the U.S. Geological Survey Hydro-Climatic Data Network) were included in the analysis, bcause modifications to streamflows could have substantial impacts on FDCs (Castellarin et al., 2013; Kroll et al., 2015). In addition to near-natural conditions, streamgages in this study have at least 40 years of daily mean streamflow records since 1950 to minimize impacts due to differences in sampling variability between sites (Vogel et al., 1998). Previous studies have focused on fitting a probability distribution to daily streamflows at small and/or intermittent streams (Croker et al., 2003; Mendicino and Senatore, 2013; Pumo et al., 2014; Rianna et al., 2011). Here, sites having an average daily flow value of zero (flows below 0.01 feet^3/second) were omitted from analysis because such intermittent sites require additional methodological considerations. These criteria resulted in 398 gages (Fig. 1) with mean daily streamflows obtained from the USGS National Water Information System (U.S. Geological Survey, 2001). Physiographic regions, which differentiate between areas of the US with similar physical and climate characteristics (Fenneman and Johnson, 1946), are also shown in Fig. 1. These regions were used to assess whether GOF metrics are related to the physiographic setting. The periods of record for the study streamgages range from 40-61 years between 1950 through 2010, and drainage areas vary from 2 to over 5,000 km^2.

4 Results

4.1 *Graphical identification of candidate distributions*

To identify candidate probability distributions, theoretical L-moment ratios are compared to sample L-moment ratios in Fig. 2a. The four-parameter kappa (KAP) distribution is represented by the area below the generalized logistic curve and above the theoretical L-moment limits. The lower bound of the five-parameter Wakeby (WAK) distribution is also plotted as a curve. Sample estimates of L-moment ratios computed from empirical FDC_{POR} at study sites are shown as points.

Empirical L-moment ratios mostly fall below the generalized logistic and generalized extreme value curves and above the Pearson type III and WAK lower bound curves (Fig. 2a). The points are clustered around the three-parameter generalized Pareto (GPA), and lognormal (LN) curves, thus these two distributions are identified as possible parent distributions. The empirical L-moment ratios are also consistent with both KAP and WAK distributions, resulting in the identification of four candidate distributions.

The clustering of points around the GPA or LN distribution curves could, in theory, be due to sampling variability. However, given a sufficiently long record, empirical L-moment ratios would be expected to fall directly on the theoretical curves if the probability distribution of daily streamflows truly arose from GPA or LN. The very large sample sizes here suggest this is unlikely, nevertheless synthetic daily streamflows were generated to test this hypothesis. The method of L-moments (Hosking and Wallis, 1997) was used to estimate distribution parameters from the ranked observed daily streamflows, or the empirical FDC_{POR}, for each study gage. FDC_{POR} led to parameters that were inconsistent with KAP at 35 sites (9%) and with WAK distributions at 244 sites (61%). Because WAK could not be fit at over half of the study streamgages, a finding encountered previously for new England (Archfield, 2009), WAK was removed from further consideration.

Based on distribution parameters for GPA, LN and KAP, data of the same record length as the daily streamflow observations at a given site were simulated and L-moment ratios computed. These synthetic L-moment ratios are plotted in Fig. 2b-d. As expected given the very large samples, the synthetic L-moment ratios for GPA and LN fall on the empirical curves representing these distributions. Thus, the scatter in L-moment ratios does not appear to be due to sampling variability, but rather reflects the complexity of the true distribution(s) from which the daily streamflows arise. Compared to GPA and LN, simulated L-moment ratios from KAP (Fig. 2d) appear more consistent with L-moment ratios estimated from empirical FDCs. Thus, KAP appears to provide the best fit among the probability distributions considered, yet given the benefits of fewer parameters in practice and the observation that some gages plot on the theoretical L-moment ratio curves, the GPA and LN hypotheses, are retained for future analyses.

4.2 *National goodness of fit comparisons*

In this section, we consider additional measures of the GOF of the GPA, LN and KAP models for approximation of FDC_{POR}. One complication involves the generation of negative streamflows, when the fitted lower bound of a distribution is less than zero. Negative streamflows were predicted at 98 sites for GPA, 159 sites for LN and 40 sites for KAP. Others have also encountered problems with the generation of negative streamflow (Archfield, 2009; Castellarin et al., 2007). To prevent infeasible negative flow predictions, distributions were constrained to ensure a theoretical lower bound of zero at study sites for which negative flows were generated. Both GPA and LN include parameters that represent the theoretical lower bounds of the distribution (Hosking and Wallis, 1997). Constraining both of these lower bound parameters to zero was relatively simple as this is equivalent to fitting two-parameter versions of GPA and LN distributions. For the KAP, the lower bound is a function of all four parameters. Therefore, enforcing a theoretical lower bound for KAP requires solving for the four parameters simultaneously while enforcing the lower bound constraint. The same approach as Castellarin et al. (2007) in constraining the KAP lower bound to zero was followed here. KAP parameters were infeasible at 42 sites (11%).

Figure 3a gives boxplots showing the range of values of LNSE across sites corresponding to the GPA, LN and KAP hypotheses. To ensure fair comparison across the three distributions, only LNSE values for sites for which KAP could be estimated (356 sites), are shown, however, the figure appears nearly identical when the additional 42 sites are included for GPA and LN. KAP shows the highest GOF, which is not surprising given that the distribution includes an additional parameter. GPA and LN both also have quite high values of LNSE (note that the y-axis ranges from 0.8 to 1). To illustrate how these LNSE values translate into GOF, example FDC_{POR} are given for three sites with varying GOF (Fig. 3b-d.) It is important to note that there was substantial variability in how FDC_{POR} appear across similar LNSE values and these are only three limited examples. First, in Fig3b, empirical and fitted FDC_{POR} with LNSE values above 0.99 for all three distributions is given. For this site, nearly the entire FDC_{POR} is captured except for the very lowest flows. A site with "good" fits, all with LNSE values between 0.93 and 0.99 is shown in in Fig3c. For this site, GPA over-estimates the highest flows and under-estimates the lowest flows. LN and KAP do a very good job with the upper tail but KAP doesn't accurately predict the lower tail. Finally, Fig3d illustrates a site where all three distributions show poor fits (LNSE values below 0.93).

To assess the magnitude of errors across exceedance probabilities, error duration curves are shown in Fig. 4. These plots illustrate how the error (represented by the ratio of predicted to observed flows) is distributed across the quantiles for GPA, LN and KAP. Values of 1 would indicate no error and above one indicates that predicted flows are greater than observed flows for a given quantile. Each grey line represents the error for a given study site given the exceedance probability. These error duration plots illustrate that there are some errors for the very highest flows (exceedances close to zero), with all three distributions dramatically over predicting the highest flows for some sites. Although far from smooth, error generally increases for the lower flows (exceedance probabilities closer to 1). This highlights the challenge of having one distribution represent the tail behavior of both low and high flows. While GPA and LN errors appear relatively comparable, errors for KAP are generally smaller across all quantiles.

4.3 *Goodness of fit by physiographic region*

Perhaps the challenges encountered above for sites with poor fits to FDC_{POR} are primarily driven by certain regions within the US. Focusing on such a large study region presents a particularly difficult challenge but also a unique opportunity to compare GOF of candidate distributions across regions within the US. Figure 5a shows boxplots of LNSE by probability distribution for the eight physiographic regions in the US which included at least 20 study sites. Sample sizes are given as well as the number of sites within each region FDC_{POR} that could not be estimated with KAP. (This appears to be a particular problem in the Piedmont region where only 8 of the 24 sites had feasible KAP parameters.) These boxplots illustrate that there are some regions in the US in which all three distributions provide a very good fit, such as new England, the Appalachian and the valley and ridge regions. Perhaps a three-parameter distribution such as GPA or LN might be adequate to describe FDCs in these region, as Fennessey (1994) found to be the case for the mid-Atlantic region. For most regions, KAP provides the best fit, which is not surprising given that it has an additional parameter compared to GPA and LN. The Cascade-Sierra mountain region does not have very high GOF for any of the three candidate distributions.

Maps of the US illustrating LNSE for GPA, LN and KAP are given in Fig. 5b. For GPA (left), nearly all "poor fits" (LNSE<0.93) are at sites in the western half of the country. Very good fits (LNSE>0.99) are found throughout the US, but are primarily clustered in new England and the mid-Atlantic regions. For LN (middle), more sites have LNSE values above 0.99 compared to GPA3, particularly in the eastern half of the country and there are fewer sites on the west coast with LNSE values below 0.93. Finally, the map of KAP LNSE (right) illustrates that, of the 356 sites which could be fit with KAP, the majority are well-approximated by KAP with LNSE values above 0.99. However, a limitation of KAP is that it could not be used to estimate FDC_{POR} at 42 sites in the study region. Martinez and Gupta (2010) found somewhat similar geographic patterns in GOF for a monthly water balance model applied across the conterminous US.

4.4 *Median annual flow duration curves*

The FDC_{POR} reflects the steady-state or long-term behavior of the frequency-magnitude relationship for streamflow. Alternatively, if one's interest is in the frequency-magnitude relationship in a typical year, median annual FDCs (FDC_{MED}) are useful (Vogel and Fennessey, 1994). Less dependent upon the specific period of record than FDC_{POR}, FDC_{MED} are increasingly applied in practice in situations which focus on hydrologic conditions for a typical year. For example, FDC_{MED} have recently been used to predict hydropower production (Mohor et al., 2015; Müller et al., 2014), evaluate regional similarity between streams under different flow conditions (Patil and Stieglitz, 2011), and characterize baseflow variability (Hamel et al., 2015). FDC_{MED} are also used to compare streamflow regimes in different catchments (Hrachowitz et al., 2009), to assess before and after watershed land-use changes (Kinoshita and Hogue, 2014) and to quantify fish passage delays

(Lang et al., 2004). More generally, FDC_{MED} are useful in testing hypotheses regarding any form of flow alteration (Kroll et al. 2015).

A few studies have attempted to fit a probability distribution to FDCs in a typical year. LeBoutillier and Wayland (1993) found a five-parameter mixed lognormal distribution to be superior to two- and three-parameter lognormal, Gamma, and generalized extreme value distributions for fitting probability distributions to mean annual FDCs of four rivers in Canada. For the mid-Atlantic U.S., Fennessey (1994) identified the GPA as a suitable distribution for both FDC_{POR} and FDC_{MED}, developed regional regression models to relate GPA model parameters to basin characteristics, and then used those models to predict FDCs at ungaged locations. FDC_{MED} can also be estimated seasonally, and seasonal FDC_{MED} have been used to evaluate impacts on ecological flow regimes (Gao et al., 2009; Lin et al., 2014; Vogel et al., 2007).

The procedure for constructing an FDC_{MED} is similar to the FDC_{POR}, but rather than ranking all recorded flows, flows are ranked within each calendar year resulting rankings 1-365 for each of the *n* years. Then, the median flow at each ranking is selected for inclusion FDC_{MED}. The majority of the FDC_{POR} and FDC_{MED} curves are generally very similar, only differing at the lowest and highest exceedance probabilities. This is because the most extreme flows on record are included in FDC_{POR}, but are not included in FDC_{MED}, as the median estimator is insensitive to outliers. See Vogel and Fennessey (1994) for a more detailed discussion of the relationship between FDC_{POR} and $FDC_{MED.}$

Figure 6a shows the relationship between empirical L-skew and L-kurtosis for FDC_{MED} at study sites. These L-moment ratios appear quite similar to those found for FDC_{POR} (Fig. 2a), as do the LNSE values shown in boxplots (Fig. 6b). As for FDC_{POR}, distributions were constrained to ensure no negative streamflows are predicted, and KAP appears to provide the best fit to $FDC_{MED.}$ Fig. 6c shows error duration curves for FDC_{MED}. The main difference with the error duration plots for FDC_{POR} (Fig. 4) is the errors are smaller at the lowest and highest exceedances (close to 0 and 1). This is not surprising given that FDC_{MED} curves are generally quite similar to FDC_{POR}, but lack the most extreme high and low flows.

5 Discussion and conclusions

Due to the complexity associated with time series of daily streamflows, the challenge set forth in this study—to identify a single probability distribution that could approximate the distribution of daily flows—was an ambitious one. Based upon multiple goodness of fit (GOF) assessments, three candidate probability distributions were identified which can approximate period-of-record (FDC_{POR}) and median annual (FDC_{MED}) flow duration curves at perennial, unregulated stream gage sites in much of the conterminous United States (US).

Many assumptions were made which should be evaluated before applying these results to a practical application, as is the case for any model. Previous work on this subject has identified the need for at least four-parameters to describe the complex distribution of daily streamflows; however, this study is unique in that the suitability of a probability distribution for streamflow is investigated at the sub-continental scale with streamgages in widely-varying physiographic and hydroclimatic settings. For these study streamgages, four-parameter kappa (KAP) was found to provide a very good fit to the distribution of daily streamflows across most of the US (at the 89% of sites with valid KAP parameters). A special case of the KAP distribution, three-parameter generalized Pareto (GPA) can provide an acceptable fit for certain regions of the US, particularly new England, Appalachian and the valley and ridge regions. Compared to GPA, three-parameter log normal (LN) was found to result in predictions with better GOF compared to GPA, particularly the pacific border and the cascade-sierra regions. To prevent the prediction of infeasible negative streamflows, all three distributions required lower bound constraints for some sites. More work on parameter estimation that enforces the conditions that observed streamflows be both non-negative and exceed theoretical distributional lower bounds is needed.

Few previous studies have sought to evaluate theoretical probability distributions for modelling FDC_{MED}, however, their growing use suggests that our findings relating to FDC_{MED} could have broad applications. We caution users of FDC_{MED} to be aware that the FDC_{MED} can only provide a window into the behavior of streamflow in a typical year, thus we recommend that whenever FDC_{MED} are used that users also illustrate the entire family of annual FDCs which gave rise to the computation of the FDC_{MED}.

There are many limitations of this work. First, daily streamflows are not independent, and thus exhibit an extremely high level of serial correlation which will impact the confidence intervals or any other form of uncertainty analysis associated with the modeled FDCs. Furthermore, daily streamflows exhibit seasonality so that they are far from being identically distributed, which is assumed whenever one attempts to fit a single distribution to a random variable. In addition, this study included only perennial and unregulated streams. While there is some existing literature on intermittent regimes (Mendicino and Senatore, 2013; Pumo et al., 2014; Rianna et al., 2011), and the impacts of human regulation on flow duration curves (Gao et al., 2009; Kroll et al., 2015), additional research on these topics would improve our understanding of flows across a wider range of streams. Finally, the seasonality of daily streamflows suggests that distributional analyses of this nature should be done at a seasonal level, as was recently carried out on a broad scale for daily precipitation (see Papalexiou and Koutsoyiannis, 2016). The definition of seasons, as well as the parent distributions which can approximate streamflows within those seasons, has been shown to vary across sites (Bowers et al., 2012). Given that gages varied over a large range of hydroclimatic conditions, a seasonal analysis was beyond the scope of this study, but we recommend that future studies consider the impact of seasonality on the GOF of FDCs. Daily streamflow varies over four or five orders of magnitude and is subject to seasonality and serial correlation. When viewed though this lens, the finding of any reasonable candidate distribution that provides some explanatory power - such as those explored here - is somewhat remarkable. Future research on intermittent sites, differences across seasons, lower bound constraints, and additional distributional types, such as mixed-distributions, should help to improve prediction of daily streamflows at ungaged sites across the globe.

Acknowledgements

We gratefully acknowledge the insightful comments of Francisco Serinaldi, Skip Vecchia, Dave Holtschlag, William Farmer, Jory Hecht and two anonymous reviewers. We are indebted to William Asquith, the author of the RStudio package *lmomco,* and J. R. M. Hosking, the author of the *lmom* package. This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant number DGE-1144081 and NSF Grant number EEC-1444926. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Archfield, S. A.: Estimation of continuous daily streamflow at ungaged locations in southern New England, PhD dissertation, Tufts University, 2009.

Archfield, S. A. and Vogel, R. M.: Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments, Water Resour. Res., 46, 1–15, doi:10.1029/2009WR008481, 2010.

Basso, S., Schirmer, M. and Botter, G.: On the emergence of heavy-tailed streamflow distributions, Adv. Water Resour., 82, 98-105, doi: 10.1016/j.advwatres.2015.04.013, 2015.

Botter, G., Zanardo, S., Porporato, A., Rodriguez-Iturbe, I. and Rinaldo, A.: Ecohydrological model of flow duration curves and annual minima, Water Resour. Res., 44(8), 1–12, doi:10.1029/2008WR006814, 2008.

Bowers, M. C., Tung, W. W. and Gao, J. B.: On the distributions of seasonal river flows:

Lognormal or power law? Water Resour. Res., 48(5), 1–12, doi:10.1029/2011WR011308, 2012.

Castellarin, A., Botter, G., Hughes, D. A., Liu, S., Ouarda, T. B. M. J., Parajka, J., Post, D. A., Sivapalan, M., Spence, C., Viglione, A. and Vogel, R. M.: Prediction of flow duration curves in ungauged basins, in Runoff prediction in ungauged basins: synthesis across processes, places and scales., edited by G. Blöschl, M. Sivapalan, and T. Wagener, pp. 135–162, Cambridge University Press. 2013.

Castellarin, A., Camorani, G. and Brath, A.: Predicting annual and long-term flow-duration curves in ungauged basins, Adv. Water Resour., 30(3), 937–953, doi: 10.1016/j.advwatres.2006.08.006, 2007.

Castellarin, A., Vogel, R. M. and Brath, A.: A stochastic index flow model of flow duration curves, Water Resour. Res., 40(3), 1–10, doi:10.1029/2003WR002524, 2004.

Doulatyari B., Betterle, A., Basso, S., Biswal, B., Schirmer, M. and Botter, G.: Predicting streamflow distributions and flow duration curves from landscape and climate, Adv. Water Resour., 83, 285-298, doi: 10.1016/j.advwatres.2015.06.013, 2015.

Fenneman, N. M., and Johnson, D, W.: Physical divisions of the United States, U.S. Geological Survey, 1:7,000,000, 1946. [online] Available from: https://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml, (last access 30 Dec 2016),

Fennessey, N. M.: A hydro-climatological model of daily streamflow for the northeast United States, PhD dissertation, Tufts University, 1994.

Fennessey, N. M. and Vogel, R. M.: Regional Flow-Duration Curves for Ungaged Sites in Massachusetts, J. Water Res. Pl.–ASCE, 116, 1–20, 1990.

Gao, Y., Vogel, R. M., Kroll, C. N., Poff, N. L. and Olden, J. D.: Development of representative indicators of hydrologic alteration, J. Hydrol., 374(1-2), 136–147, doi:10.1016/j.jhydrol.2009.06.009, 2009.

Hamel, P., Daly, E. and Fletcher, T. D.: Which baseflow metrics should be used in assessing flow regimes of urban streams? Hydrol. Process, doi:10.1002/hyp.10475, 2015.

Hosking, J. R. M.: L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, J. R. Stat. Soc., 52, 105–124, 1990.

Hosking, J. R. M.: The four-parameter kappa distribution, IBM J. Res. Dev., 38, 1–8, 1994.

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

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U. and Fenicia, F.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrolog. Sci. J., 58(6), 1198-1255, doi: 10.1080/02626667.2013.803183, 2013.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Dawson, J. J. C., Dunn, S. M. and Malcolm, I. A.: Using long-term data sets to understand transit times in contrasting headwater catchments, J. Hydrol., 367(3), 237–248, doi:10.1016/j.jhydrol.2009.01.001, 2009.

Hughes, D. A. and Smakhtin, V.: Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves, Hydrolog. Sci. J., 41(6), 851–871, doi:10.1080/02626669609491555, 1996.

Kinoshita, A. M. and Hogue, T. S.: Increased dry season water yield in burned watersheds in Southern California, Environ. Res. Lett., 10(1), 1–9, doi:10.1088/1748-9326/10/1/014003, 2014.

Klemeš, V.: Tall tales about tails of hydrological distributions. I, J Hydrol Eng. 5(3), 227-3, doi: 10.1061/(ASCE)1084-0699(2000)5:3(227), 2000.

Kroll, C. N., Croteau, K. E. and Vogel, R. M.: Hypothesis tests for hydrologic alteration, J. Hydrol., 530, 117-126, doi: 10.1016/j.jhydrol.2015.09.057, 2015.

Lang, M., Love, M. and Trush, W.: Improving stream crossings for fish passage, National Marine Fisheries Report. 2004.

LeBoutillier, D. W. and Waylen, P. R.: A Stochastic Model of Flow Duration Curves, Water Resour. Res., 29, 3535–3541, doi: 10.1029/93WR01409, 1993.

Li, M., Shao, Q., Zhang, L. and Chiew, F.H.: A new regionalization approach and its application to predict flow duration curve in ungauged basins, J. Hydrol.,389(1), 137-145, doi: 10.1016/j.jhydrol.2010.05.039, 2010.

Lin, K., Lian, Y., Chen, X. and Lu, F.: Changes in runoff and eco-flow in the Dongjiang River of the Pearl River Basin, China, Front. Earth Sci., 8(4), 547–557, doi:10.1007/s11707-

014-0434-y, 2014.

Martinez, G. F. and Gupta, H. V.: Toward improved identification of hydrological models: A diagnostic evaluation of the “abcd” monthly water balance model for the conterminous United States, Water Resour. Res., 46(8), doi:10.1029/2009WR008294, 2010.

Mendicino, G. and Senatore, A.: Evaluation of parametric and statistical approaches for the regionalization of flow duration curves in intermittent regimes, J. Hydrol., 480, 19–32, doi:10.1016/j.jhydrol.2012.12.017, 2013.

Mohor, G. S., Rodriguez, D. A., Tomasella, J. and Júnior, J. L. S.: Exploratory analyses for the assessment of climate change impacts on the energy production in an Amazon run-of-river hydropower plant, J. Hydrol.: Regional Studies, 4, 41–59, doi:10.1016/j.ejrh.2015.04.003, 2015.

Müller, M. F., Dralle, D. N. and Thompson, S. E.: Analytical model for flow duration curves in seasonally dry climates, Water Resour. Res., 50(7), 5510–5531, doi:10.1002/2014WR015301, 2014.

Müller, M. F., and Thompson, S. E.: Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes, Hydrol. Earth Syst. Sci., 20(2), 669-83, doi:10.5194/hess-20-669-2016, 2016.

Papalexiou, S. M. and Koutsoyiannis D: A global survey on the seasonal variation of the marginal distribution of daily precipitation. Adv. Water Resour., 94,131-45, 2016

Patil, S. and Stieglitz, M.: Hydrologic similarity among catchments under variable flow conditions, Hydrol. Earth Syst. Sci., 15(3), 989–997, doi:10.5194/hess-15-989-2011, 2011.

Pumo, D., Viola, F., La Loggia, G. and Noto, L. V.: Annual flow duration curves assessment in ephemeral small basins, J. Hydrol., 519(PA), 258–270, doi:10.1016/j.jhydrol.2014.07.024, 2014.

Rianna, M., Russo, F. and Napolitano, F.: Stochastic index model for intermittent regimes: from preliminary analysis to regionalisation, Nat. Hazards Earth Syst. Sci., 11, 1189-1203, doi:10.5194/nhess-11-1189-2011, 2011.

Schaefli B., Rinaldo, A. and Botter, G.: Analytic probability distributions for snow-dominated streamflow, Water Resour. Res., 49 (5), 2701-2713, doi: 10.1002/wrcr.20234, 2013.

Segura, C., Lazzati, D. and Sankarasubramanian, A.: The use of broken power-laws to describe the distributions of daily flow above the mean annual flow across the conterminous U.S, J. Hydrol., 505(C), 35–46, doi:10.1016/j.jhydrol.2013.09.016, 2013.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrolog. Sci. J., 48(6), 857–880, doi:10.1623/hysj.48.6.857.51421, 2003.

Smakhtin, V. U.: Low flow hydrology: a review. J. Hydrol., 240(3):147-86, doi: 10.1016/S0022-1694(00)00340-1, 2001.

Smakhtin, V.Y., Masse, B.: Continuous daily hydrograph simulation using duration curves of a precipitation index, Hydrol. Process, 14, 1083–1100, doi: 10.1002/(SICI)1099-1085(20000430)14:6<1083::AID-HYP998>3.0.CO; 2000.

Stedinger, J. R., Vogel, R. M. and Foufoula-Georigiou, E.: Frequency analysis of extreme events, in Handbook of Hydrology, edited by D. Maidment, pp. 1–69, McGraw-Hill. 1993.

U.S. Geological Survey: *Hydro-Climatic Data Network (HCDN-2009)*, [online] Available from: http://water.usgs.gov/osw/hcdn-2009/, (last access 15 August 2015), 2015.

U.S. Geological Survey: *National Water Information System*, [online] Available from: http://waterdata.usgs.gov/nwis/, (last access 15 August 2015), 2015.

Viola, F., Noto, L.V., Cannarozzo, M.A.R.C.E.L.L.A. and Loggia, G.L.: Regional flow duration curves for ungauged sites in Sicily. Hydrol. Earth Syst. Sci., 15(1), 323-331, 2011

Vogel, R. M. and Fennessey, N. M.: L moment Diagrams should replace product moment diagrams, Water Resour. Res., 29, 1745–1752, 1993.

Vogel, R. M. and Fennessey, N. M.: Flow Duration Curves I: New Interpretation and Confidence Intervals, J. Water Res. Pl. -ASCE, 120, 485–504, 1994.

Vogel, R. M. and Fennessey, N. M.: Flow Duration Curves II, Water Resour. Bull., 31, 1029–

1039, 1995.

Vogel, R. M., Sieber, J., Archfield, S. A., Smith, M. P., Apse, C. D. and Huber-Lee, A.: Relations among storage, yield, and instream flow, Water Resour. Res., 43(5), doi:10.1029/2006WR005226, 2007.

Vogel, R. M., Tsai, Y. and Limbrunner, J. F.: The regional persistence and variability of annual streamflow in the United States, Water Resour. Res., 34, 3445–3459, 1998.

Yokoo, Y. and Sivapalan, M.: Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis, Hydrol. Earth Syst. Sci., 15(9), 2805–2819, doi:10.5194/hess-15-2805-2011, 2011.

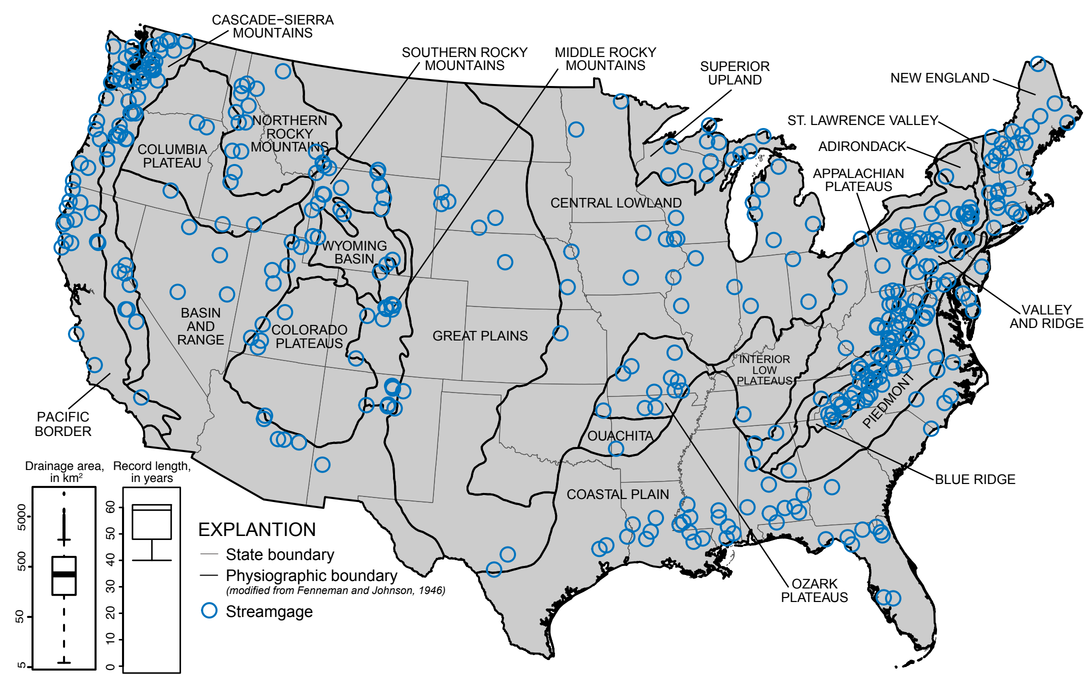

Figure 1. Map of the conterminous United States showing physiographic regions and the stream gages included in the study. Boxplots on the lower left show the range of drainage areas and record lengths represented by study stream gages.

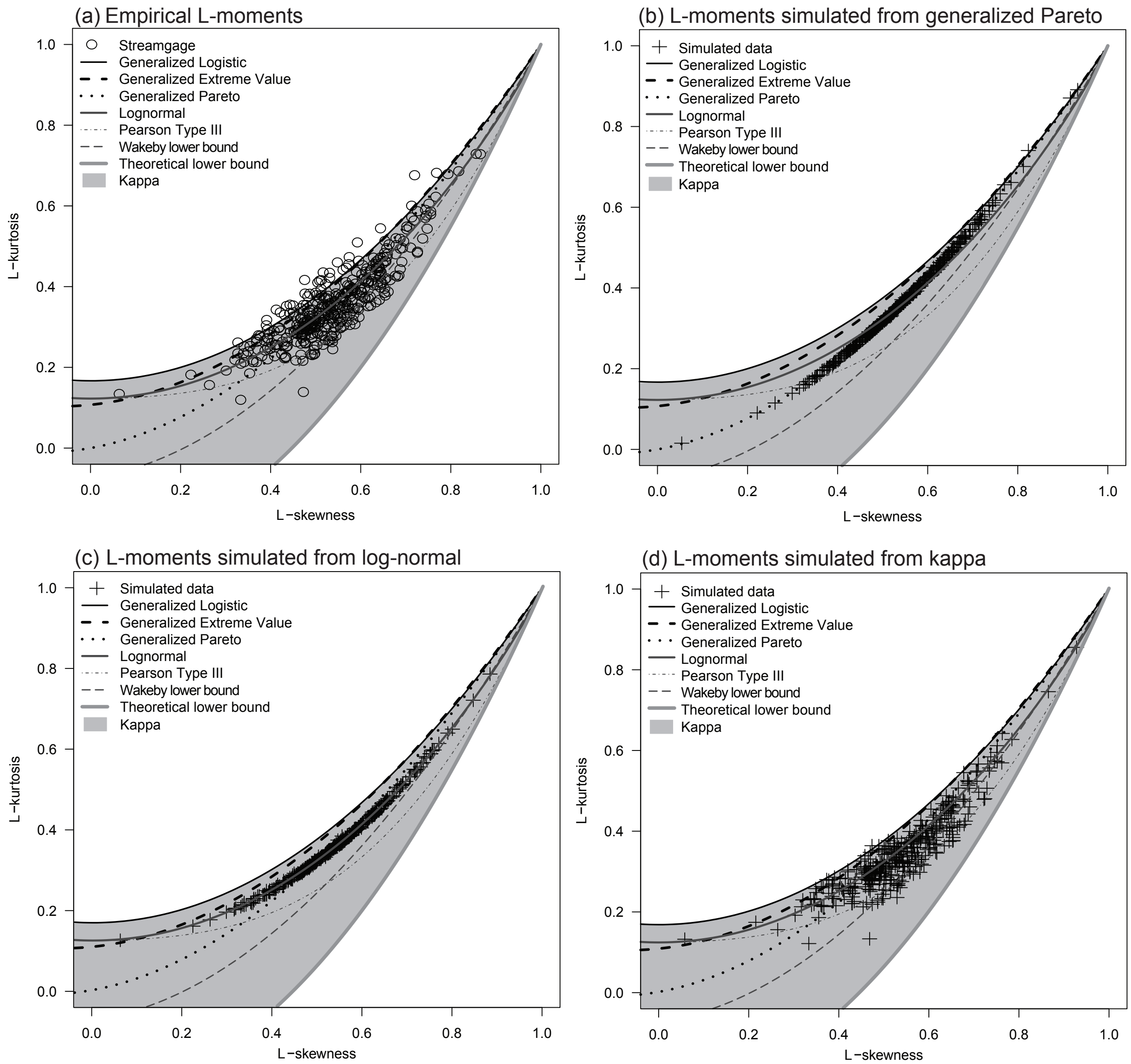

Figure 2. L-moment diagrams for (a) daily streamflows and flows simulated from (b) generalized Pareto, (c) lognormal and (d) kappa distributions.

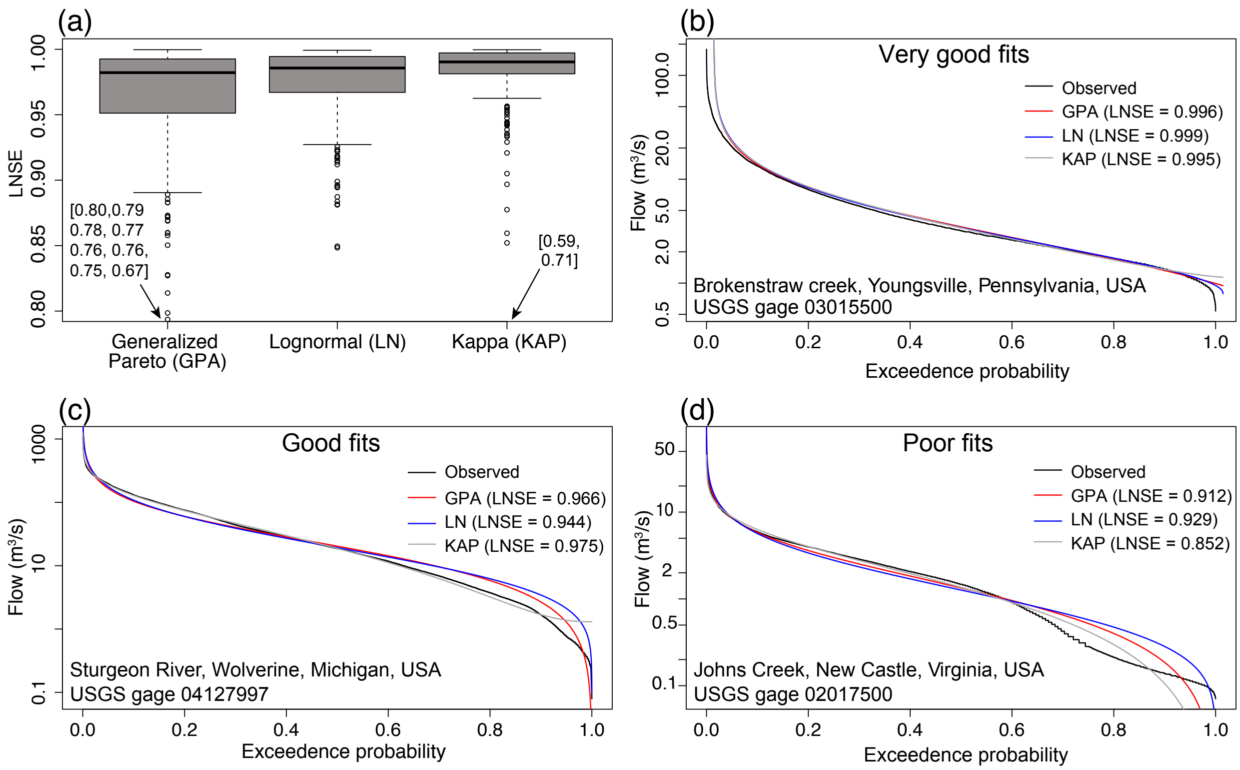

Figure 3. (a) Boxplots showing the range of stream gage Nash-Sutcliffe efficiencies for natural logarithms of daily streamflows (LNSE) based on hypothesized generalized Pareto, lognormal and kappa distributions; and example stream gage sites with (b) very good fits (LNSE above 0.99); (c) good fits (LNSE between 0.93 and 0.99); and (d) poor fits (LNSEs below 0.93).

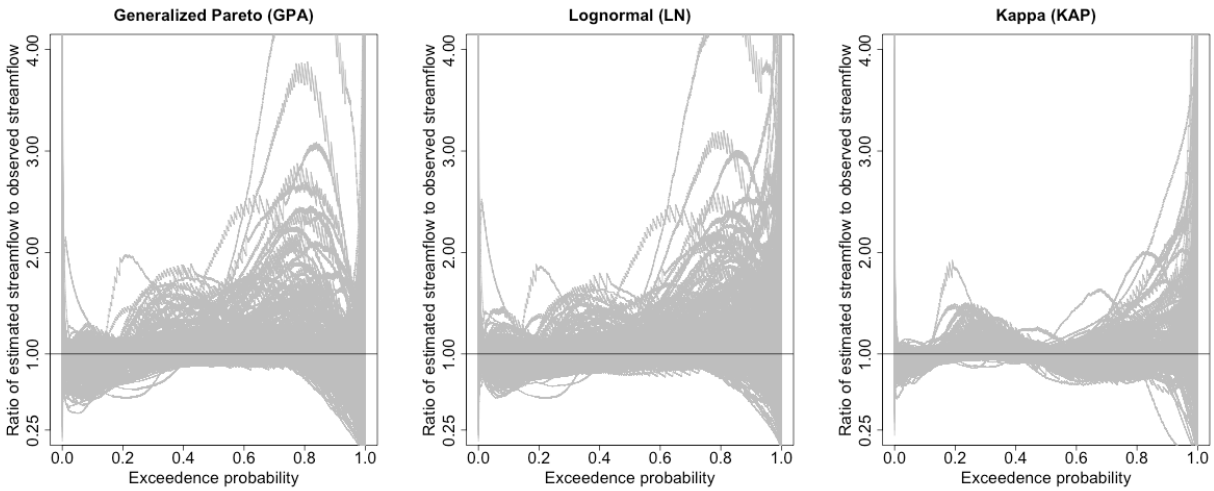

Figure 4. Error duration plots illustrating the range of errors (ratio of estimated to observed streamflows) across exceedance probabilities for generalized Pareto, lognormal and kappa hypotheses. Each grey line represents the estimated relative error for a study stream gage and the black horizontal line at 1 shows a benchmark for no error.

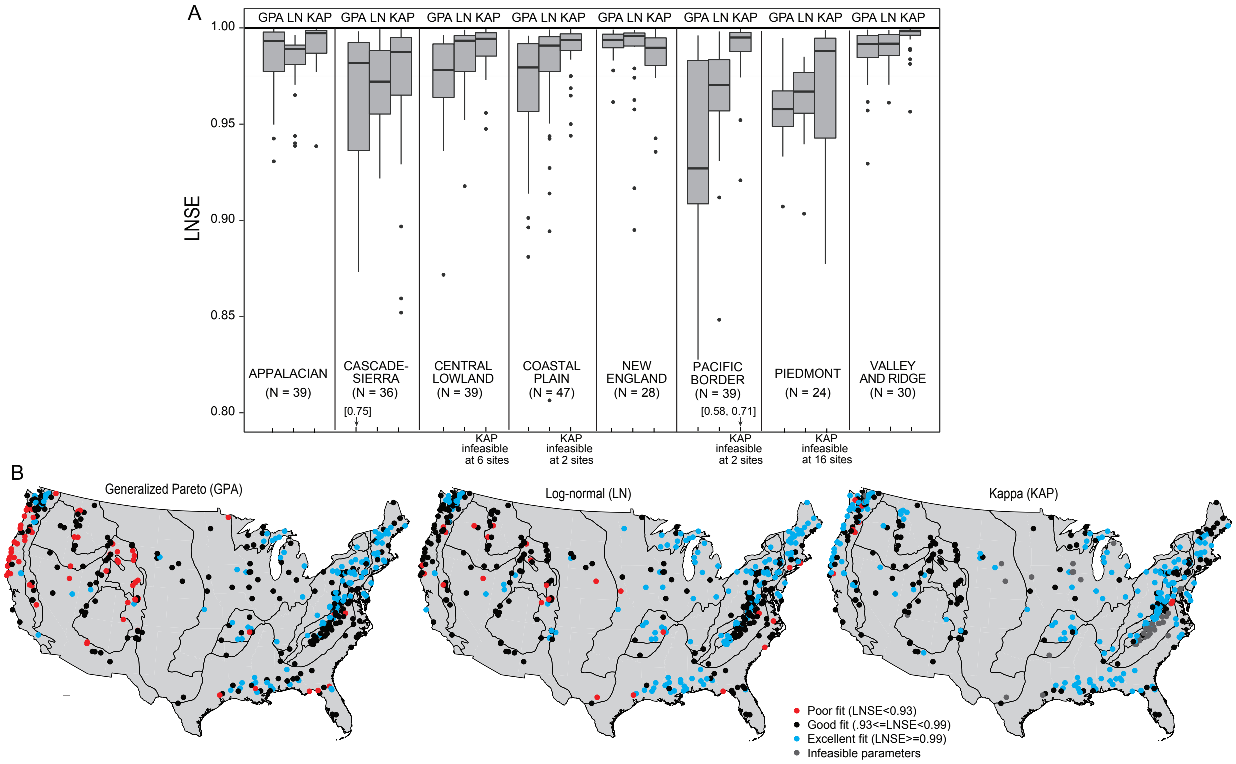

Figure 5. (a) By physiographic region, boxplots of stream gage Nash-Sutcliffe efficiencies of natural logarithms of daily streamflows (LNSE) based on hypothesized generalized Pareto (GPA), lognormal (LN) and kappa (KAP) distributions. Below the region name, the number of study gages located in that region is listed as *N*. Only regions with at least 20 study gages are shown to facilitate relatively fair comparisons across regions. (b) Maps of the conterminous United States illustrating stream gage LNSE values for GPA, LN and KAP hypotheses.

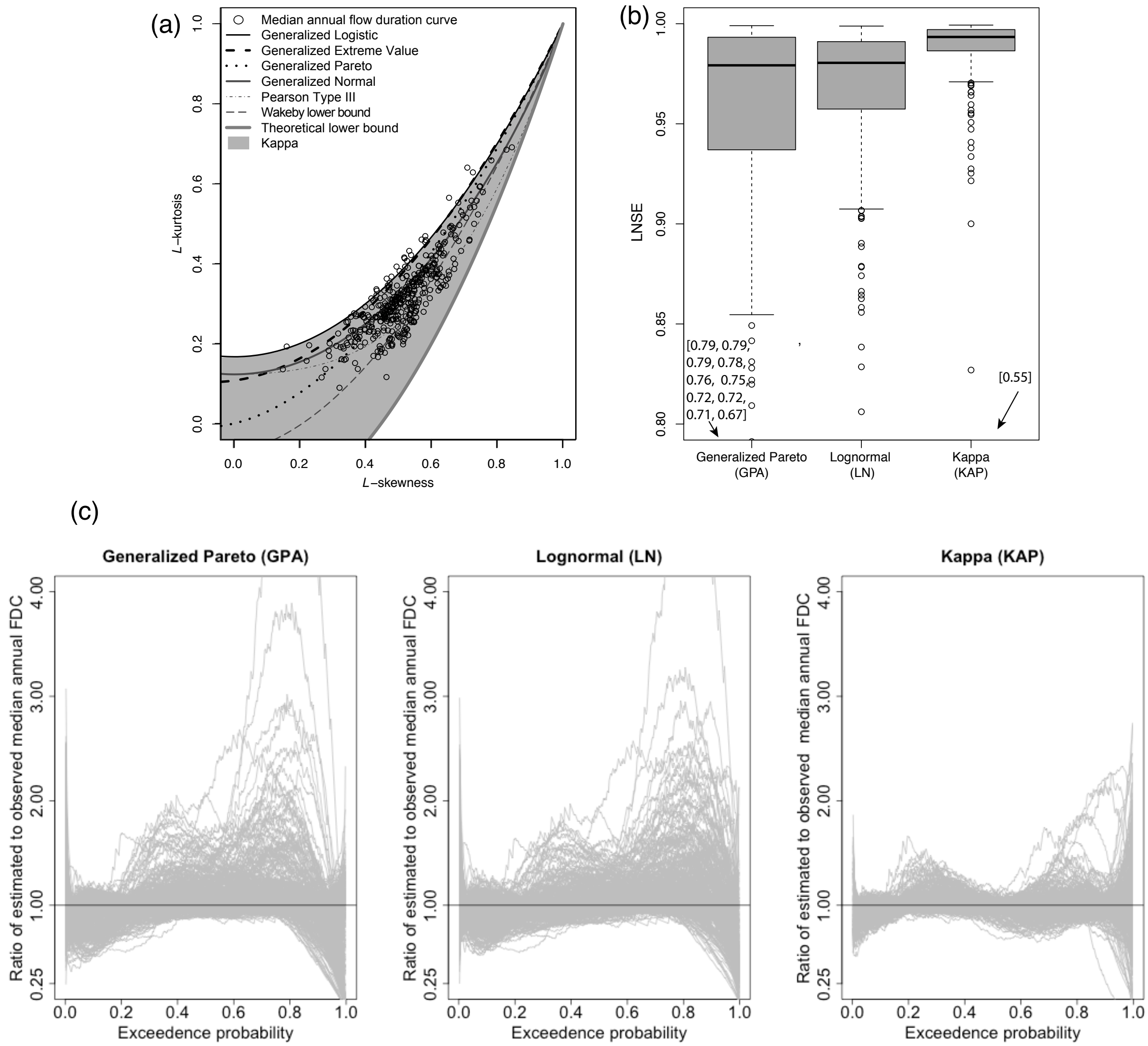

Figure 6. (a) L-moment diagram with empirical L-moment ratios of the median annual flow duration curves (FDC_{MED}) estimated at study stream gages; (b) boxplots of stream gage Nash-Sutcliffe efficiencies for natural logarithms of FDC_{MED} (LNSE) based on hypothesized generalized Pareto (GPA), lognormal (LN) and kappa (KAP) distributions; (c) Error duration plots for FDC_{MED} illustrating the range of errors (ratio of estimated to observed FDC_{MED}) across exceedance probabilities for GPA, LN and KAP hypotheses. Each grey line represents the estimated relative error for a study stream gage and the black horizontal line at 1 shows a benchmark for no error.