1 **Statistical distributions for monthly aggregations of precipitation and**

2 **streamflow in drought indicator applications**

3

4 Cecilia Svensson[1],*, Jamie Hannaford[1], Ilaria Prosdocimi[1, 2]

5

6 [1] Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford,

7 Wallingford, Oxfordshire, OX10 8BB, UK

8 [2] Now at the Department of Mathematical Sciences, University of Bath, Claverton Down,

9 Bath, Somerset, BA2 7AY, UK

10 * Corresponding author, email: csve@ceh.ac.uk

11

12 15 September 2016

13 **Running title:** Statistical distributions for precipitation and streamflow

14

15 **Key points**

16 The suitability of twelve statistical distributions for characterizing long-duration precipitation

17 and streamflow data was evaluated

18 The novel and flexible three-parameter Tweedie distribution has a lower bound at zero and

19 describes the data well

20 The Tweedie distribution is advocated for calculation of precipitation and streamflow drought

21 indices

22

23

Abstract

Drought indicators are used as triggers for action and so are the foundation of drought monitoring and early warning. The computation of drought indicators like the standardized precipitation index (SPI) and standardized streamflow index (SSI) require a statistical probability distribution to be fitted to the observed data. Both precipitation and streamflow have a lower bound at zero, and their empirical distributions tend to have positive skewness. For deriving the SPI, the Gamma distribution has therefore often been a natural choice. The concept of the SSI is newer and there is no consensus regarding distribution. In the present study, twelve different probability distributions are fitted to streamflow and catchment average precipitation for four durations (1, 3, 6, and 12 months), for 121 catchments throughout the United Kingdom. The more flexible three- and four-parameter distributions generally do not have a lower bound at zero, and hence may attach some probability to values below zero. As a result, there is a censoring of the possible values of the calculated SPIs and SSIs. This can be avoided by using one of the bounded distributions, such as the reasonably flexible three-parameter Tweedie distribution, which has a lower bound (and potentially mass) at zero. The Tweedie distribution has only recently been applied to precipitation data, and only for a few sites. We find it fits both precipitation and streamflow data nearly as well as the best of the traditionally used three-parameter distributions, and should improve the accuracy of drought indices used for monitoring and early warning.

Index terms

1816 Estimation and forecasting

1854 Precipitation

1860 Streamflow

1804 Catchment

Keywords

Standardized precipitation index, standardized streamflow index, statistical distribution, Tweedie distribution, drought index, United Kingdom.

1. Introduction

It is widely written that drought is a complex phenomenon that is notoriously hard to define. The challenges of drought definition have been acknowledged for at least 30 years, certainly since *Wilhite and Glantz* [1985], and recently *Lloyd-Hughes* [2014] argued that a universal definition is simply impracticable. Given the vagueness of the concept of drought, it is perhaps unsurprising that there has been a proliferation of drought indicators over the last two decades: *Lloyd-Hughes* [2014] counted over 100 different indicators, with no sign of abatement in this trend in the international literature. While there is some merit in applying different indicators for different purposes – recognizing the widely acknowledged multi-faceted nature of the drought hazard, from meteorological, hydrological, agricultural, socio-economic, etc – there is also benefit in seeking indicators that are consistent and comparable. Indicators provide crucial information to support the recognition of drought onset and termination, and thus duration, as well as the relative severity and rarity of events. They are used as triggers for action and communication (or declaration of drought) and so are the foundation of drought monitoring and early warning.

One of the more popular indicators, now recommended by the World Meteorological Organization (WMO) [*Hayes et al.*, 2011] as an indicator for meteorological drought, is the Standardized Precipitation Index (SPI) first formulated by *McKee et al.* [1993], see section 3.4 for details. A key advantage of the SPI is that, by virtue of the standardization process, it enables comparison between locations and between different times of year. It is also flexible in that it can be aggregated across a range of timescales (e.g. 1, 3, 6, 12 months), which are relevant for different types of drought impacts. As such, the SPI has been widely applied in monitoring and early warning systems (e.g. see the review by *Bachmair et al.*, 2016) and is extensively used for other purposes.

While the SPI is widely seen as a de facto indicator of choice, it is not without limitations. By construction the SPI is inherently sensitive to the time period used for standardization, and the choice of statistical distribution used to effect the transformation to a normal distribution. The impact of record length and standardization period has been analyzed [e.g. *Wu et al.* 2005, *Núñez et al.*, 2014]. The choice of distributions has been discussed since the early days of the SPI; the question is far from academic and can have major implications for drought severity and characteristics. The original paper [*McKee et al.*, 1993] employed the gamma distribution, and this has been widely adopted and is sometimes seen as a default; it has

85 generally been shown to perform relatively well across most of Europe [*Lloyd-Hughes and*

86 *Saunders*, 2002]. However, a later systematic test across Europe [*Stagge et al.*, 2015] found

87 that while the Gamma distribution could still broadly be recommended as a default, other

88 distributions perform better in some areas, and at some aggregation timescales. Other authors

89 have noted limitations with the Gamma distribution, in particular under (over) estimation of

90 wet (dry) extremes [e.g. *Sienz et al.*, 2012]. Some authors [e.g. *Guttman*, 1999] have

91 advocated three parameter distributions such as the Pearson Type 3 as more flexible

92 alternatives.

93 In the last decade there has been an increased interest in the application of the SPI concept to

94 other domains of the hydrological cycle. A Standardized Precipitation and

95 Evapotranspiration Index (SPEI), developed by *Vicente-Serrano et al.* [2010] has become

96 very popular, and other studies have sought to apply standardization to runoff [*Shukla and*

97 *Wood*, 2008], streamflow [e.g. *Vicente-Serrano et al.* 2012], groundwater [*Bloomfield and*

98 *Marchant*, 2013] and snowmelt [*Staudinger et al.*, 2014]. Finding suitable distributions is as

99 important for these variables as it is for the SPI, but this topic has been relatively under-

100 researched. Unlike for the SPI and SPEI, for streamflow there is not even a broad consensus

101 on the most appropriate distributions to use. For the standardized streamflow index (SSI),

102 there are only a few published studies (to the authors' knowledge) that address choice of

103 distribution, and they all tend to focus on a limited number of sites or a localized

104 geographical domain. *Vicente-Serrano et al.* [2012] examined distributions for 98 sites in the

105 Ebro basin in Spain, finding that no distribution worked best, partly due to the substantial

106 variations in streamflow regimes, even over a relatively limited area. These authors instead

107 advocate an approach of empirically choosing best distributions for each month and location.

108 *Soláková et al.* [2014] examined a range of distributions for a single site in Italy, compared to

109 a non-parametric approach. For groundwater, the Standardized Groundwater Index (SGI)

110 [*Bloomfield and Marchant*, 2013] has been developed using a non-parametric approach due

111 to the inappropriateness of most distribution fits for a range of boreholes in the UK. For

112 meteorological data covering larger areas, most of the systematic tests of distributions have

113 been applied to readily-available, continuous gridded datasets, e.g. *Stagge et al.* [2015] for

114 Europe, rather than to catchment observations.

115 The aim of this paper is to systematically test distributions for application to the SPI and SSI

116 using a large dataset of 121 near-natural catchments across the whole of the UK, with respect

117 to

118   1) goodness-of-fit and general distributional behavior, including visual inspection of the

119   fitted probability density functions,

120   2) the degree to which unbounded distributions have mass below zero (as hydrological

121   data have a lower bound at zero),

122   3) effects of seasonality and catchment characteristics on the goodness-of-fit.

123 One immediate purpose of the study is to provide guidance for application of such

124 standardized indicators in the UK; the SPI has not been routinely employed beyond a few

125 research studies [e.g. *Hannaford et al.* 2011; *Bloomfield and Marchant*, 2013; *Folland et al.*

126 2015], with the SSI even less so. Standardized indices are not routinely used in drought

127 monitoring and early warning (with the exception of in Scotland, where an SPI variant is used

128 [*Gosling*, 2014]), but there is increasingly an appetite for consistent indicators for drought

129 definition [e.g. *Collins et al.*, 2015]. There is a pressing need to tackle this question, given the

130 diversity of catchment types seen in the UK, to provide recommendations for application of

131 the SPI and SSI in monitoring and early warning systems and other applications. A related

132 paper [*Barker et al.*, 2016] further tests the utility of the SPI and SSI in the UK, using them to

133 investigate the propagation from meteorological to hydrological drought and relating the

134 propagation characteristics to catchment characteristics. By virtue of the scale of the dataset

135 and diversity of climatic and landscape characteristics seen in the UK (as discussed in

136 *Barker et al.*, [2016]), we envisage the test application presented here could be of wide

137 international relevance, given the paucity of studies that have addressed comprehensively the

138 question of distribution choice for both precipitation and river flow in parallel.

139

140 **2. Data**

141 2.1 Precipitation and streamflow data

142 Monthly mean streamflow and precipitation data for 121 catchments in the UK benchmark

143 network were used for the analysis. The benchmark network comprises catchments with near-

144 natural river flow regimes and long, good-quality flow records [*Bradford and Marsh*, 2003].

145 Here, "near-natural" is defined as catchments having minimal net influence of abstractions,

146 discharges or impoundments. The benchmark catchments are located throughout the country

147 (Figure 1a), and can be considered to be representative of the range of catchment conditions

148 encountered across the UK. The catchment areas vary between 3.07 and 4587 km^2 (Figure

149   1b). The selected catchments have at least 30 years of both river flow and precipitation

150   records within the period October 1929 to September 2014 (Figures 1c and 1d).

151   Daily mean streamflow and catchment average monthly precipitation totals were provided by

152   the National River Flow Archive (NRFA) [*Marsh and Hannaford*, 2008]. For streamflow,

153   each individual month was required to have at least 25 days of valid daily mean flow

154   observations for a monthly mean flow (expressed in m^3/s) to be calculated and used in this

155   study. The mean and median record lengths are 47.3 and 44.9 years, respectively, when

156   missing data are included. Almost half of the catchments (59 of 121) have complete records

157   between their individual start and end dates, and the remainder has on average 2.7% missing

158   data. The longest record, for the Dee at Woodend in northeast Scotland, spans the whole 85-

159   year period October 1929 to September 2014 and has no missing data. There were two

160   ephemeral streams in the initial catchment selection, with 4% and 13% of monthly flows

161   equal to zero. Because of this relatively large amount of zeroes in the dataset, these

162   catchments were removed so that all monthly flow observations for the selected 121

163   catchments exceed zero. This was done to avoid the complication of treating the data for

164   some catchments differently to the rest, and to enable a straightforward comparison of the

165   proposed distribution functions. It also avoids having to distinguish between instances of a

166   zero flow recording because of a lack of rainfall, and because of frozen conditions (although

167   the latter tend not to be long-lasting in the UK).

168   The periods of record for the catchment average monthly precipitation data differ slightly

169   from those for the streamflow records, with mean and median record lengths being slightly

170   longer, at 47.8 and 46.6 years, respectively. However, the precipitation records for all

171   catchments end in December 2012, nearly two years before the last river flow record. The

172   longest record is again the one for the Dee at Woodend, starting in October 1929 and ending

173   in December 2012, without any missing data. Only four of the catchments have missing

174   precipitation data between their individual start and end dates, on average 2.3%. The monthly

175   precipitation totals equaled zero for one month each in four catchments. These single

176   observations were set to 0.01 to avoid having zeroes in the dataset. The monthly precipitation

177   totals were converted to monthly averages expressed in mm/day.

178

179

180

181   2.2 Catchment descriptors

182   Indices describing the characteristics of each catchment were also supplied by the NRFA

183   [*Marsh and Hannaford*, 2008; *Bayliss*, 1999]. These include catchment AREA, SAAR (the

184   standard average annual rainfall 1961-1990), SMDBAR (the average soil moisture deficit

185   1960-1990. This is only available for 31 catchments), and BFIHOST, a base flow index

186   which is a measure of catchment responsiveness derived using the 29-class Hydrology of Soil

187   Types (HOST) classification. There is a strong association between BFIHOST and the

188   Baseflow Index derived using a hydrograph separation approach, where the Baseflow Index

189   is a measure of how large the groundwater contribution to streamflow is. Both indices vary

190   between 0 and 1, with higher values indicating a greater contribution of base flow and

191   therefore higher catchment storage, or permeability. Here we consider catchments having

192   BFIHOST values exceeding 0.5 to be permeable and catchments with values less than 0.5 to

193   be impermeable. Most of the permeable catchments are located in the southeast of the UK

194   (see Figure 1a for aquifer outcrop areas).

195

196   **3. Methods**

197   3.1 Overview

198   Data series were extracted for each calendar month separately, so that, for example, a series

199   could consist of river flows for January 1965, January 1966, January 1967, and so on. These

200   data series were then transformed to standardized indices: the standardized precipitation

201   index (SPI) for precipitation data and, separately, the standardized streamflow index (SSI) for

202   river flow data. The first step in the transformation was to fit a statistical probability

203   distribution to the data series (e.g. to the series of January river flows). A range of different

204   distributions considered potentially suitable for precipitation and river flow data were tried in

205   this step. Next, the fitted distribution was used to transform the original data to the quantiles

206   of a Normal distribution with mean = 0 and standard deviation = 1. These quantiles, or

207   "normal scores" of the Normal distribution constitute the standardized indices. A range of

208   goodness-of-fit tests were then applied to the standardized indices to investigate which

209   statistical distribution best fitted the original precipitation and river flow data. The novel part

210   of the present study is to systematically test a range of statistical distributions for suitability

211   of describing precipitation and river flow data. In addition to the goodness-of-fit tests, the

212   mass below zero was assessed for the original distributions, and the influences of seasonality

213   and catchment characteristics on the goodness-of-fits were investigated.

214

215   3.2 Aggregated time series

216   Separate data series were derived for each calendar month, so that series consisting of data

217   for all Januaries were extracted separately from series consisting of data for all Februaries,

218   etc. In addition to these monthly series, longer aggregations of 3, 6 and 12 months were also

219   derived, with separate (backward-looking) aggregated series ending in each of the 12

220   calendar months of the year. The aggregations are averages of the monthly values in the

221   aggregation duration, so that, for example, the 6-month precipitation aggregation for August

222   2010 is the average daily rainfall total for March to August 2010. A missing value in one or

223   more of the constituent months of the aggregation leads to a missing value for the total

224   aggregation.

225

226   3.3 Statistical distribution functions

227   In total twelve different probability distributions with two, three or four parameters were

228   considered to be potentially suitable for fitting to the monthly aggregations of precipitation

229   and river flow data (Table 1). Not all of the distributions have a lower bound at zero (like

230   observed precipitation and streamflow do), but unbounded distributions were included in the

231   study as they have been used for deriving standardized indices in the past [e.g. *Guttman*,

232   1999]. The magnitude of the associated problem of part of the fitted distribution potentially

233   falling below zero is therefore also assessed in the present study.

234   The sole four-parameter distribution investigated is the Kappa distribution, and the three-

235   parameter distributions include the Generalized Logistic, Pearson type 3, Generalized

236   Extreme Value (GEV) and Tweedie. Of these, only the Tweedie distribution has a lower

237   bound at zero. The two-parameter distributions include the Gamma, Lognormal, Normal,

238   Gumbel and Weibull distributions. Of these, the Normal and Gumbel distributions lack a

239   lower bound. Probably with the exception of the Tweedie, these are all well-known

240   distributions to the hydrologist, and apart from the Tweedie [e.g. *Dunn and Smyth*, 2005] and

241   the Kappa distributions [e.g. *Hosking and Wallis*, 1997], their probability density functions

242   are described in *Stagge et al.* [2015].

243   In addition, a lower bound at zero was imposed on the three- and four-parameter distributions

244   for which a bounded version of the distribution is not already investigated. That is, a lower

245   bound at zero was imposed for the Generalized Logistic and the Kappa distributions. When

246   such a constraint is applied to the Pearson type 3 distribution, it reduces to the Gamma

247   distribution, which is already included. Similarly, the already included 2-parameter Weibull

248   distribution is a variation of the GEV with a lower bound at zero. The imposition of the lower

249   bound reduces the number of parameters to be estimated by one, so that the three- and four-

250   parameter distributions become two- and three-parameter distributions, respectively. In

251   figures and tables of this paper, the short-hands "GenLog", "P3", "Kappa3" and "Kappa4"

252   may be used to denote, in turn, the Generalized Logistic, the Pearson type 3, and the 3- and 4-

253   parameter Kappa distributions.

254   To the authors' knowledge the Tweedie distribution has only rarely been applied to

255   hydrological data. *Dunn* [2004] and *Hasan and Dunn* [2011] applied it to monthly rainfalls in

256   Australia, and found that it fitted the data well. The Tweedie is in fact a family of

257   distributions, whose special cases include the Normal ($\xi = 0$), Poisson ($\xi = 1$), Gamma

258   ($\xi = 2$), and Inverse Gaussian ($\xi = 3$) distributions [e.g. *Tweedie*, 1981; *Dunn and Smyth*,

259   2005]. When the parameter ξ has a value between 1 and 2, the distribution describes series

260   that can contain exact zeroes as well as positive continuous data, whereas for $\xi \geq 2$ the

261   support is for positive continuous data only [*Hasan and Dunn*, 2011]. Although the present

262   study does not include data equal to zero, the Tweedie family of distributions has two

263   attractive features: the lower bound (for $\xi \geq 1$), and that its three parameters make it flexible.

264   Of the other proposed distributions, only the Kappa distribution with a lower bound imposed

265   at zero have both these characteristics. However, a drawback of the Tweedie is that the

266   probability density functions can only be written in closed form for the special cases where ξ

267   equals 0, 1, 2 or 3. For other values of ξ numerical methods need to be used, which makes

268   parameter estimation more time-consuming [e.g. *Dunn and Smyth*, 2008].

269   Please see Appendix A for further details about the Tweedie distribution, and Appendix B for

270   details about the distribution fitting procedures.

271

272   3.4 Transformation to the standardized indices, SPI and SSI

273   Once distributions are fitted, the observations (i.e. the aggregated series) are transformed to a

274   Normal distribution with mean = 0 and standard deviation = 1, N(0,1), to obtain the

275   standardized index as described by, for example, *McKee et al.* [1993], *Guttman*, [1999], and

276   *Lloyd-Hughes and Saunders* [2002]. That is, each observation corresponds to a quantile, *x*, of

277   the fitted cumulative distribution function, $F(x) = P(X \leq x)$, of the distribution of choice (say,

278   for example, the Gamma distribution). By setting

279    $F(x) = G(y),$

280   where $G(y)$ is the cumulative distribution function of the N(0,1) distribution, the

281   corresponding quantile *y* of the N(0,1) distribution can be found for each observation. The

282   quantile *y*, or "normal score", is the unitless standardized index for the observation, i.e. the

283   SPI or SSI. Hence, the SSI is derived in the same manner as the SPI, but using streamflow

284   rather than precipitation data.

285   During dryer than average conditions the index will be negative and during wetter than

286   average conditions the index will be positive. Because of uncertainty at the extremes of the

287   distribution, estimates of the index exceeding an absolute value of 5 were truncated and set to

288   the limiting value. Approximately 95% of the calculated indices will occur within the

289   range -2 to +2, and 68% within the range -1 to +1. The exact definitions of different drought

290   severities differ between studies, but spells in the latter range are often denoted as normal to

291   mildly dry/wet, and spells outside the range -2 to +2 as extremely dry/wet.

292

293   3.5 Testing the goodness-of-fit

294   A range of goodness-of-fit tests were applied to investigate how well the different statistical

295   distributions describe the observed data. However, the tests were applied to the transformed

296   indices, rather than to the original precipitation and river flow data. In this way, all the

297   goodness-of-fit tests are applied to normally distributed data with mean = 0 and standard

298   deviation = 1, and the results are straightforward to compare.

299   The tests include the Shapiro-Wilk test, the Anderson-Darling test, the Cramér-von Mises test

300   and the Kolmogorov-Smirnov test. The Anderson-Darling test is a modification of the

301   Cramér-von Mises test, giving more weight to the tails of the distribution [e. g. *Farrell and*

302   *Rogers-Stewart*, 2006]. *Razali and Wah* [2011] compared the power of several tests, and

303   concluded that of the four tests investigated the Shapiro-Wilk test was the most powerful test

304   for normality closely followed by the Anderson-Darling test, and that the Lilliefors test

305   always outperformed the least powerful Kolmogorov-Smirnov test. However, none of these

306   tests perform well for small sample sizes of fewer than 30 cases [*Razali and Wah*, 2011],

307   which is the smallest number of years of record required for a catchment to be included in the

308   present study. Because the Shapiro-Wilk, Anderson-Darling and Cramér-von Mises tests

309   gave nearly identical results in this study (and the Kolmogorov-Smirnov test accepted nearly

310   all of the fitted three-parameter distributions so that no distinction between them could be

311   made), only results from the Shapiro-Wilk test will be presented. The significance level

312   chosen for this study is 5% (p-value = 0.05 for the Shapiro-Wilk test).

313

314   3.6 Visual inspection of fitted distributions

315   A visual inspection of a selection of the fitted distributions was carried out, mainly to assist in

316   the interpretation of why certain distributions were rejected, but also to investigate if there

317   were any obvious systematic differences in how the distributions fit to the data. For the 1-

318   month precipitation and streamflow aggregations, all the instances for which the p-value from

319   the Shapiro-Wilk test of normality was <0.05 were plotted as histograms, and the curve for

320   the rejected probability density function added. On each plot, a selection of other fitted

321   distributions were plotted for comparison with the rejected one. To look for clear systematic

322   differences, each selection of distributions was plotted regardless of aggregation duration and

323   Shapiro-Wilk p-value, and subjected to a brief visual overview.

324

325   3.7 Proportion of the fitted distribution below zero

326   Probability distributions that are not bounded below at zero may have a proportion of the

327   fitted distribution occurring for sub-zero quantiles, that is, the cumulative probability

328   distribution $F(x=0) > 0$. Precipitation and river flow data, on the other hand, cannot be

329   negative and in the present study $F(x=0)$ should be zero as all the data exceed zero. The

330   magnitudes of $F(x=0)$ for the fitted distributions were therefore assessed.

331

332   3.8 Effects of seasonality and catchment characteristics

333   The effects of seasonality and catchment characteristics on the goodness-of-fit were

334   investigated for two different measures of goodness-of-fit. First, the rejection rate for each

335   distribution was used, which means stratifying the p-values from the Shapiro-Wilk test on

336 values below and above 0.05 (the significance level). Second, the average p-values from the

337 Shapiro-Wilk test were investigated. Rejection rates were compared for the winter and

338 summer seasons, and for permeable versus impermeable catchments. Seasonal average p-

339 values for each catchment were used in a correlation analysis with different catchment

340 descriptors. Because of the non-normal distribution of some of the catchment descriptors,

341 correlations were estimated using the non-parametric Spearman's rho. Again, a 5%

342 significance level was employed, based on a two-sided test.

343 The winter season here comprises October to March and the summer season April to

344 September. Only the 1-month aggregation was used for the seasonal analysis, as the longer

345 the aggregation period for the standardized index is, the less clear the distinction between the

346 seasons become.

347

348 **4. Results**

349 4.1 Goodness-of-fit

350 The goodness-of-fit of the selected statistical distributions to the standardized precipitation

351 and streamflow data was assessed, and the results are presented in Table 2. The proportion of

352 occurrences for which the normal distribution cannot be rejected at the 5% significance level

353 (p-value = 0.05) by the Shapiro-Wilk test is shown for all four aggregation durations

354 separately, as well as for all the durations lumped together. In all cases, the 12 calendar

355 months and 121 catchments, are lumped together. Table 2 only shows results for distributions

356 with rejection rates less than 15% for SPI and 20% for SSI, and the remainder of the paper

357 will concern only these distributions. Of the two-parameter distributions, only the Gamma is

358 included in the table. This means that the only distributions with a lower bound at zero that

359 remain in the study are the Gamma, the Tweedie and the three-parameter Kappa distributions.

360 All the unbounded three- and four-parameter distributions are retained. Overall, the rejection

361 rates are higher for the streamflow than for the precipitation data. This seems reasonable

362 since streamflow is affected by the natural and human-induced state of the catchment, e.g.

363 soil moisture content and ground water levels, which can lead to highly variable and non-

364 linear responses to the, in comparison, more regularly-behaving precipitation data.

365 In the past, the Gamma distribution was often used when calculating the SPI [e.g. *McKee et*

366 *al.*, 1993; *Lloyd-Hughes and Saunders*, 2002; *Stagge et al.*, 2015] because it has a lower

367 bound at zero, which agrees with the observed distribution. Others [e.g. *Guttman*, 1999] have

368 argued that three-parameter distributions such as the Pearson type 3 provide a better fit.

369 Although the Gamma distribution has a positive skewness, which generally reflects the

370 behavior of the observed data, distributions with separate location, scale and skewness

371 parameters are more flexible. Table 2 shows that the Gamma distribution is rejected

372 considerably more frequently than any of the three- or four-parameter distributions for both

373 precipitation (9.9%) and streamflow (19.2%).

374 For both the SPI and the SSI, the distributions rejected the fewest times are the four-

375 parameter and three-parameter Kappa distributions. For the SPI, the Pearson type 3 has the

376 same rejection rate as the three-parameter Kappa distribution, at 1.2% of cases. The Pearson

377 type 3 and the Kappa distributions perform consistently well across all the different durations

378 (Table 2), whereas the rejection rates vary for the other distributions. The Gamma,

379 Generalized Logistic and the two Kappa distributions tend to be rejected less frequently for

380 longer durations than shorter, for both SPI and SSI (Table 2), but the variation with duration

381 is not necessarily monotonic and some distributions seem to fit best for the intermediate

382 durations.

383 For streamflow, the one month aggregation period is likely to be more useful for drought

384 characterization than it is for precipitation, as there is already a degree of aggregation

385 inherent in streamflow time series, primarily resulting from catchment storage (e.g. *Chiverton*

386 *et al.*, 2015). For the 1-month SSI, the Gamma distribution is rejected for 29.5% of the cases,

387 whereas the four-parameter Kappa and the Tweedie show the best fits with rejection rates of

388 3.1% and 5.2%, respectively (Table 2).

389

390 4.2 Visual inspection of the fitted distributions

391 Histograms of the data and fitted distributions were plotted and visually inspected. Two

392 different sets of comparisons are presented for the 1-month aggregations. Figures 2a-c and

393 3a-c) show examples of the comparison of the fitted Tweedie and Kappa distributions.

394 Figures 2d-f and 3d-f show examples of the comparison between the Gamma, the Tweedie

395 and the best-fitting (according to the Shapiro-Wilk test) of the remaining three-parameter

396 distributions for the SPI and SSI, respectively. The Gamma was included because of its

397 traditional use for the SPI, and the Tweedie because it too is bounded below at zero. This

398 meant the Gamma, Tweedie and Pearson type 3 distributions were compared with each other

399   for the precipitation data (Figures 2d-f), and the Gamma, Tweedie, Generalized Logistic

400   (better than the GEV overall) and GEV (better than the Generalized Logistic for the 1-month

401   aggregation) distributions were compared for the streamflow data (examples are shown in

402   Figures 3d-f). Table S1 summarizes the overall findings of the present study for these

403   distributions.

404   With the exception of the very flexible Kappa distributions, the tested theoretical

405   distributions generally struggle to fit bimodal or multimodal data (Figures 2 and 3).

406   However, for non-trivial proportions of cases, both the three- and four-parameter Kappa

407   distributions show signs of over-fitting, resulting in sharp drops/rises and/or sharp bounds at

408   the lower and/or upper ends of the distribution, as well as the peak of the probability density

409   function occurring in unexpected locations (Figures 2a-c and 3a-c). In addition, the four-

410   parameter Kappa distribution does not necessarily encompass all the observed data within its

411   lower and upper bounds (Figures 2a, 2c and 3a), although this occurs also for other

412   distributions such as the GEV (Figure 3f). The effect on the SPI and SSI of observed data

413   occurring outside the bounds will be that the index value becomes undefined (or infinite).

414   The fitted distribution dropping/rising sharply at the tails means that the magnitudes of

415   droughts (or wet periods) will not be well distinguished by the SPI and SSI. The Kappa

416   distributions generally provide a better fit than the Tweedie distribution, as measured by the

417   Shapiro-Wilk p-value (Table 2, Figure 4). However, from an SPI and SSI application

418   perspective, the Tweedie characterizes the tails of the distributions better (Figures 2a-c and

419   3a-c).

420   The two-parameter Gamma distribution often gets rejected for bimodal and multimodal data,

421   whereas most of the time the "traditional" three-parameter distributions, although struggling,

422   are accepted (e.g. Figures 2d and 3d). These more flexible distributions tend to fit a

423   reasonably symmetric probability density function, with the Tweedie distribution fitting a

424   somewhat more (positively) skewed curve than the Generalized Logistic, GEV and Pearson

425   type 3 distributions. This "inbetween" behavior explains why the fit of the accepted Tweedie

426   distributions is often slightly worse than that of the more commonly used three-parameter

427   distributions. This can be seen in the generally lower average of the p-values from the

428   Shapiro-Wilk test for the Tweedie distribution compared with the other distributions (Figure

429   4). For precipitation data, the instances for which the Gamma and Tweedie distributions are

430   not accepted typically coincides with the fitted Pearson type 3 curve having a slight or

431   pronounced negative skewness (similar to Figure 2d). For streamflow, the typical situation

432   where the Gamma and Tweedie are rejected is for very peaked, positively skewed data,

433   particularly for cases with one or more very extreme values (e.g. Figure 3e). The less flexible

434   fit of the Gamma distribution also sometimes leads to the lower tail not describing the data

435   well (Figure 2f).

436   In about half of the cases where the Pearson type 3 distributions are not accepted for the 1-

437   month precipitation data, a sharply peaked curve with a very positive skewness has been

438   fitted (e.g. Figure 2b). The fits of the Tweedie and Gamma distributions tend to be both less

439   peaked and less skewed in these cases, and are generally accepted. A closer investigation of

440   the rejected Pearson type 3 occurrences suggests that, particularly for bimodal and

441   multimodal data, the maximum likelihood method has converged on sets of sub-optimal

442   parameter values, as the curves plotted using parameter values based on L-moments provide a

443   (generally accepted) fit more similar to the Tweedie and Gamma distributions. For the

444   Generalized Logistic and GEV fitted to 1-month streamflow data, the distributions fitted

445   using L-moments also often differ substantially to those fitted using maximum likelihood.

446   However, in these cases the data often have a single peak combined with extreme outliers

447   (similar to Figure 3e) and the generally less peaked density curves based on the L-moments

448   do not on the whole provide a better fit as measured by the p-value from the Shapiro-Wilk

449   test.

450   In very nearly all the cases where the Generalized Logistic distribution fitted to streamflow

451   data is rejected, the fitted density curve is more peaked than for the GEV, Tweedie and

452   Gamma distributions. In about a third of the cases the distribution of the observed data is *very*

453   peaked, similar to Figure 3e. However, in the example in Figure 3e the fitted Generalized

454   Logistic distribution is accepted, and in fact, the Generalized Logistic seems to cope well

455   with very peaked data compared with these other distributions, particularly when also very

456   extreme values are present in the dataset (e.g. Figures 3e-f). The Gamma distribution copes

457   particularly poorly with very extreme values.

458   Apart from the above mentioned general difficulties in fitting distributions to very peaked or

459   multi-peaked data, or to series that contain very extreme values, it is difficult to find any

460   particular patterns for the rejected GEV distributions. However, the GEV is so flexible that

461   the fitted distributions not infrequently have an abrupt upper bound when fitted to negatively

462   skewed data (Figure 3f). Correspondingly, there is a problem when the fitted distribution has

463   a lower bound greater than zero, which can occur for the four-parameter Kappa (Figures 2a-c

464   and 3a), the Generalized Logistic (Figure 3e), the Pearson type 3 (Figures 2e-f), as well as for

465   the GEV (Figure 3e and 5b).

466   Abrupt upper bounds, as well as lower bounds greater than zero, can be avoided altogether by

467   using the Tweedie distribution, which similarly to the Gamma and three-parameter Kappa

468   distribution is bounded below at zero but not above. However, as discussed above, the

469   Tweedie allows for more flexibility in the skewness compared with the Gamma. Table 2

470   shows that the fit of the Tweedie is second only to the four-parameter Kappa distribution for

471   1-month streamflow, and is very nearly as good as the Generalized Logistic for all the

472   aggregation durations lumped together (but both are outperformed by the two Kappa

473   distributions). When lumping all the durations together for precipitation, the Tweedie is

474   outperformed by all but the Gamma and GEV distributions, despite having a very low

475   rejection rate at 2.4%.

476

477   4.3 Behavior at the lower tail

478   Figures 2e, 3b and 3e show how the fitted Pearson type 3, Generalized Logistic, GEV and

479   four-parameter Kappa distributions can have part of their mass for quantiles below zero (i.e.

480    $F(0) > 0$), whereas neither precipitation totals or streamflows can be negative. Observed data

481   could be zero, but in the present study all the data are above zero.

482   The Tweedie distribution is bounded below, but in contrast to the Gamma and three-

483   parameter Kappa distributions, both of which have zero probability for zero

484   precipitation/streamflow, certain parameter values of the Tweedie distribution lead to a

485   positive mass at zero. This means that potentially, if the fit is such that $F(0) \gg 0$, then the

486   same problem would occur as for the distributions that are not bounded below at zero. That

487   is, there would be a lower limit to the SPIs and SSIs. However, for the catchments in the

488   present study, the fitted Tweedie distributions reflect the observed data well and do not have

489   much mass at zero. As can be seen from Table 3, the vast majority of cases has $F(0) < 0.01$,

490   corresponding to an SPI/SSI of -2.33.

491   Some of the other distributions can also be considered to have acceptably low proportions of

492    $F(0)$, particularly for precipitation. For example, the Pearson type 3 has $F(0) \leq 0.03$

493   (corresponding to an SPI of -1.88) for all but one out of in total 5808 cases. For streamflow,

494   the values of $F(0)$ are slightly larger.

16

Figure 5 shows how the fit of the density curve at the lower tail can affect the estimated SPI and SSI during a drought. The 1976 drought was long and severe in the southeast of the UK. Figure 5 shows example histograms and fitted distributions for the Tove at Cappenham Bridge (hydrometric number 33018). For streamflow, the Gamma distribution fits particularly poorly to the data at the lower tail, and data never reach into the lower tail of the distribution. This leads to underestimation of the drought, in the same way as when three-parameter distributions have part of their mass for quantiles less than zero, i.e. $F(0)>0$ (Figures 5b and 5e). The fitted GEV has a lower bound, leading to an overestimation of drought severity (Figures 5b and 5e). The smallest flow observation is below the lower bound of the GEV, and the SSI 1 – which therefore originally was estimated at minus infinity – was capped at a value of -5.

For precipitation, the fitted probability density curves added to the histograms of the observed 6-month aggregation show no obvious visual differences (Figure 5a), but the time series plots reveal how the different distributions' behaviors at the lower tail affect the SPI 1 and SPI 6 estimates of severe droughts (Figures 5c and 5d).

The SPI 6 time series shows a similar drought progression as the SSI 1 during 1975 and 1976, suggesting that a 6-month aggregation of precipitation reflects the storage in the catchment. However, note that streamflow (SSI 1) in this particular catchment also responds to heavy 1-month precipitation (SPI 1) occurrences not seen in the longer SPI 6 aggregation. This reflects the non-linear relationship between precipitation and runoff, and supports the need for a separate drought index for streamflow. The geology of the Tove catchment is predominantly Chalk, but with 50% overlain by Boulder Clay, a low-permeability superficial deposit. This accounts for the slow response (storage of water in the permeable Chalk aquifer) as well as the fast response (fast runoff from the impermeable areas overlain by clay).

4.4 Effect of seasonality and catchment characteristics on goodness-of-fit

For 1-month precipitation, all distributions except the four-parameter Kappa had smaller rejection rates for winter than for summer, but the differences between the seasons were small. In absolute terms, all the differences were smaller than 1.2 percentage units, except for the Gamma which had a 2.5 percentage unit difference. For 1-month streamflow, all the distributions had smaller rejection rates for winter than for summer. The differences were

527   smaller than 4.4 percentage units for all the distributions, except for the Gamma and the

528   Pearson type 3, which had 22.9 and 28.9 percentage unit difference, respectively.

529   For all the distributions except the Gamma, the absolute difference in rejection rate (lumping

530   all the aggregation durations together) between streamflow in permeable and impermeable

531   catchments was 1.6 percentage units or less, with no consistent variation in the direction of

532   change. For the Gamma distribution, the difference was 6.8 percentage units, with fewer

533   rejections for the impermeable catchments. When only the 1-month aggregation is

534   considered, the differences become slightly larger but still vary in the direction of change.

535   The absolute differences are smaller than 4.0 percentage units, except for the Gamma

536   distribution for which the difference is 13.9 percentage units with fewer rejections in

537   impermeable catchments.

538   Figure 6 shows the seasonal average p-value from the Shapiro-Wilk test for 1-month

539   streamflow in each catchment, for the same selection of distributions as used for the visual

540   inspection in section 4.2. In Figure 6, each dot represents the average of six p-values

541   (averages over April – September for summer, and over October – March for winter). The

542   Gamma distribution is the only distribution for which the average p-values are smaller than

543   0.05 – the statistical significance level for rejection – and these low values all occur in

544   summer. The other distributions also tend to have lower, and spatially more variable, average

545   p-values in summer than in winter, echoing the findings above that rejection rates are higher

546   in summer than in winter. However, the fact that the average p-values are generally much

547   greater than 0.05 suggests that when one of the better-fitting distributions is rejected it tends

548   to be an isolated occurrence, rather than reflecting a larger-scale failure over several months

549   in the same season. The corresponding maps for the 1-month precipitation analysis did not

550   reveal any obvious seasonal or geographical differences in the goodness-of-fits (not shown).

551   Correlations between seasonal average p-values and catchment descriptors are shown in

552   Table 4. There is little evidence of catchment characteristics (AREA and SAAR) influencing

553   the goodness-of-fit of the 1-month precipitation data, particularly in winter, as most

554   distributions fit these data well. For 1-month streamflow, the majority of the correlations are

555   rather weak and not significant at the 5% level, although a couple of correlations each for the

556   Gamma and GEV distributions exceed 0.4. Broadly, correlations are often positive (but not

557   necessarily significant) for SAAR and AREA, and negative for SMDBAR. Correlations with

558   BFIHOST are also mostly negative, but with notable exceptions for the Generalized Logistic

559   and GEV distributions in summer.

560

561   4.5 Influence of season and duration on the Tweedie parameter ξ

562   As mentioned in the methods section, the Tweedie is a family of distributions whose special

563   cases include the Normal ($\xi = 0$), Poisson ($\xi = 1$), Gamma ($\xi = 2$), and Inverse Gaussian

564   ($\xi = 3$) distributions [e.g. *Tweedie*, 1981; *Dunn and Smyth*, 2005]. The value of the parameter

565   ξ is constrained in the estimation procedure to be at least 1.2 (see Appendix B for further

566   details), which means that the estimated Tweedie distribution will never correspond to a

567   Normal or a Poisson distribution. However, in the height of winter (January) the ξ values tend

568   to approach 1 from above for both 1-month precipitation and 1-month streamflow (not

569   shown), in response to the distributions being rather symmetric. In contrast, more positively

570   skewed distributions in the height of summer (July), lead to ξ values being clustered between

571   2 and 3, i.e. suggesting distributions more similar to the Gamma and Inverse Gaussian. Most

572   of the time ξ values do not go far above 3 for 1-month precipitation, whereas for 1-month

573   streamflow in the summer half-year they are not unusual up to 5 and may go as high as 7.

574   Distributions tend to become more symmetric with increasing aggregation duration, and the

575   longer the duration the more the ξ values tend to approach 1 from above.

576

577   **5. Discussion**

578   Following the general acceptance of the SPI as a tool for drought monitoring, there has been

579   a growing interest in the application of standardization methods to other variables such as

580   streamflow. Time series of drought indices like the SPI and SSI require a statistical

581   distribution function to be fitted to the observed (generally monthly) precipitation and

582   streamflow data. Previous research has underlined the challenges associated with finding

583   appropriate probability distributions for these purposes, and there is generally limited

584   consensus on which to choose, which highlights the importance of testing distributions in the

585   region of interest. For example, *Stagge et al.* [2015] tested a range of distributions for the

586   SPI, using gridded precipitation data across Europe. Although they concluded that the

587   Gamma distribution generally showed a good fit for aggregations longer than one month, the

588   UK was one of the regions for which rejection rates were high. The present paper uses

19

589 observed streamflows and catchment average rainfalls from a diverse set of catchments in the

590 UK. It extends the SPI study of *Stagge et al.* [2015] by also investigating the Generalized

591 Logistic, Pearson type 3, GEV, Tweedie and four-parameter Kappa distribution, as well as

592 versions of the Generalized Logistic and Kappa distributions constrained to have a lower

593 bound at zero (making them two- and three-parameter distributions, respectively). In

594 addition, the inclusion of distributions that do not necessarily have a lower bound meant we

595 also investigated how many of the fitted distributions have a significant proportion of their

596 mass below zero.

597 Longer aggregations of data tend to have a more symmetric shape of the distribution than

598 shorter aggregations, which often have a positive skewness. Different distributions therefore

599 fit the data with varying success, and the practitioner could apply different distributions to

600 different aggregation durations and types of data. The ideal statistical distribution would be

601 bounded below by zero, but would have enough flexibility to fit the full range of behaviors

602 exhibited by the observed data for UK conditions. The Shapiro-Wilk test for goodness-of-fit

603 would suggest that the three-parameter Kappa distribution fulfils these expectations.

604 However, both the three- and four-parameter Kappa distributions were found to over-fit the

605 data, and to include fitted probability density functions with abrupt drops/rises and, in the

606 case of the four-parameter Kappa, sharp bounds at its lower and upper tails (Figures 2a-c and

607 3a-c). Such behavior at the tails is highly detrimental for accurately discerning the severity of

608 extreme droughts (or periods of wetness) as measured by the SPI or SSI. Further, when the

609 peak of the probability density function is displaced (often resulting in a negative skewness,

610 as in Figures 2a-c and 3c), this also results in a bias in the general characterization of dry

611 versus wet conditions. Typically, conditions that should be considered wet are labelled dry.

612 The over-fitting is a result of the Kappa distribution being very flexible, and the sample sizes

613 being rather small and the data exhibiting bi- or multimodality. While increased sample sizes

614 can be expected with time, at present it may be better to use a less flexible distribution like

615 the Tweedie. Future work on the three-parameter Kappa distribution could include

616 investigating the parameter values that lead to sharp drops at the tails of the probability

617 density function and then, if possible, constraining the parameter space accordingly when

618 fitting the distribution.

619 Of the options investigated in the present study, the Tweedie family of distributions largely

620 meets the criteria of a sensible and reasonably flexible fit combined with a lower bound at

621   zero. It can therefore be expected that future applications of SPI and SSI in the UK would

622   benefit from employing this distribution. Given the suitability of the Tweedie distribution

623   across a wide range of catchment types, it is likely that it will perform well also in other

624   settings. Given the previous lack of suitability of any one parametric method in past SSI

625   applications [*Vicente Serrano et al.*, 2012; *Soláková et al.*, 2014], we recommend further

626   research to examine the suitability of the Tweedie distribution in other environments. Its

627   general suitability for both SPI and SSI suggests it could also be applied to standardized

628   indicators used for other compartments of the hydrological cycle. The fact that the Tweedie

629   distribution can have positive mass at zero makes it particularly suitable for arid and semi-

630   arid climates and for ephemeral streams, which is desirable given the acknowledged issues

631   associated with application of the SPI in arid locations [*Wu et al.*, 2007].

632   Despite its clear potential, there are practical drawbacks associated with the Tweedie

633   distribution. Only for a small number of special cases can the distributions be written in

634   closed form. Numerical methods therefore need to be used, making distribution fitting time

635   consuming. Thus, while the Tweedie family is the most suitable choice when computational

636   resources allow, for some applications it may be less desirable than conventional

637   distributions. For example, for large gridded datasets, computer time requirements are likely

638   to be high. However, this does not preclude its use, as once distribution parameters are

639   derived, future applications (e.g. in updating a monitoring and early warning system on a

640   monthly basis using parameters already estimated for a fixed reference period) would be no

641   more computationally demanding than for other distributions.

642   This study has also provided valuable information on the applications of other distribution

643   functions. Although not ideal, for the catchments and aggregation durations used in the

644   present study, the proportions of the fitted theoretical distributions falling below zero may be

645   acceptably small for some of the best-fitting three-parameter distributions: the Pearson type 3

646   (for precipitation), and the GEV and Generalized Logistic (for streamflow) distributions.

647   When investigating statistical distributions suitable for the standardized precipitation

648   evapotranspiration index (SPEI), *Vicente-Serrano et al.* [2010] found that goodness-of-fit

649   tests struggled to distinguish between candidate distributions. Instead, they made the final

650   selection based on the behavior of the tails of the distributions. They found that the thicker

651   tail of the Generalized Logistic (termed Log-logistic in their paper) could more easily

652   accommodate the extremes of the lower tail, than could the Pearson type 3, the Lognormal

653 and the GEV distributions. *Vicente-Serrano and Beguería* [2015] investigated the difference

654 between the GEV and the Generalized Logistic in even more detail regarding goodness-of-fit,

655 behavior at the tails of the distribution, and fraction of cases with no solution, and reaffirmed

656 their preference for the Generalized Logistic. However, *Stagge et al.* [2016] subsequently

657 argued that the GEV models the extreme tails better and that the Generalized Logistic

658 consistently underestimate these values, although they conclude that there is significant

659 uncertainty due to extrapolation regardless of distribution. Considering the subjectivity of the

660 visual inspection in the present study and the limited sample sizes used for the goodness-of-

661 fit tests, we will not argue either way from a goodness-of-fit perspective of the extreme tail.

662 However, a thicker tail like that of the Generalized Logistic distribution is an advantage when

663 parameters have been estimated for a fixed reference period, and the same distribution

664 parameters are later used when updating the series [compare *Tanguy et al.*, 2015]. The

665 reference period may not have included very extreme values, and a thick-tailed distribution

666 makes it less likely that an unprecedentedly large future observation results in a standardized

667 index that is undefined or equal to infinity. In this context it is also worth mentioning that an

668 imposed lower bound needs to be at zero, rather than above (or below) zero. A lower bound

669 above zero will result in calculated standardized climate indices that are either undefined or

670 equal to negative infinity for observations that fall below the lower bound. Note that using a

671 fixed reference period assumes that data is stationary, and the distribution parameters would

672 need to be re-assessed to take into account the effect of climate, or other environmental,

673 change.

674 The rejection rates for the distributions tended to be greater in summer than in winter for both

675 precipitation and streamflow, although the differences were generally rather small for

676 distributions that fitted the data well. For both variables, the seasonal differences presumably

677 reflect the more erratic behavior of convective activity in summer, and the more regular

678 behavior of frontal precipitation in winter. For streamflow, soil moisture deficits are also

679 more variable in summer than in winter, leading to a less regular runoff response. These

680 results can be compared with those of *Stagge et al.* [2015], who found no seasonal

681 differences in rejection rates for Weibull distributions fitted to 1-month gridded precipitation

682 data across Europe.

683 Rather than stratifying on overall rejection/acceptance rates, using the seasonal average p-

684 value from the Shapiro-Wilk test for individual catchments makes possible a spatial

685 evaluation based on a more nuanced measure of goodness-of-fit. For streamflow (Figure 6),

686   the most spatially and seasonally consistent (least variable) p-values are those for the

687   Tweedie and unconstrained Kappa distributions, although for summer both struggle with an

688   area of small, and predominantly very permeable, catchments in the southeast. Permeable

689   catchments with a large groundwater contribution to flows exhibit persistence of wet and dry

690   spells that can last several years [e.g. *Wilby et al.*, 2015]. This can lead to clusters of high

691   and/or low values in the record (and thus bi- or multi-modal empirical distributions), making

692   statistical distribution-fitting difficult. Compared with the northwest, the southeast has a

693   comparatively continental climate: evaporation is larger, which leads to higher soil moisture

694   deficits and potentially erratic flow responses, particularly in permeable catchments. Also,

695   although average rainfall is lower, heavy convective rainfalls in summer are more common.

696   For small catchments, this means that localized convective rainfalls can have a greater impact

697   on the streamflow than it would in larger catchments. The correlation analysis between

698   summer average p-values and catchment characteristics supports the above reasoning in so

699   far as to the sign of the correlations, and the pattern can largely be generalized to most of the

700   distributions, and to winter. However, statistically significant relationships are few and

701   mainly occur for the poorer-fitting distributions.

702

703   **6. Conclusions**

704   Drought indicators can be used as triggers for action and declaration of drought, and form the

705   foundation of drought monitoring and early warning. Standardized indices such as the SPI

706   (for precipitation) and SSI (for streamflow) allow fair comparison between different locations

707   and different seasons. They are also flexible in that they can be aggregated across a range of

708   timescales (e.g. 1, 3, 6, 12 months), which are relevant for different types of drought impacts.

709   The derivation of the SPI and SSI require a statistical distribution to be fitted to the observed

710   data. Here, we have presented a study of the goodness-of-fit and tail behavior of twelve

711   different distributions for a diverse set of 121 catchments in the UK, as well as the influence

712   of seasonality and catchment characteristics on these fits.

713   Both precipitation and streamflow data are bounded below at zero, as precipitation and flows

714   cannot be negative. Their empirical distributions also tend to have positive skewness, and

715   therefore the Gamma distribution has often been a natural and suitable choice for describing

716   the data statistically [e.g. *Stagge et al.*, 2015]. However, after transformation of the data to

23

717 Normal distributions to obtain the SPIs and SSIs for the UK catchments, the distributions are

718 rejected in 9.9% and 19.2% of cases, respectively, by the Shapiro-Wilk test.

719 The best fitting distributions are the three- and four-parameter Kappa distributions (the

720 former being a version of the latter with a lower bound at zero imposed). However, both were

721 seen to over-fit to the data, resulting in sharp drops and rises of the probability density

722 function near the tails, making it difficult to accurately discern the severity of extreme

723 droughts (or periods of wetness) from the calculated SPI and SSI.

724 Other three-parameter distributions traditionally used in hydrological applications, such as

725 the Pearson type 3 for precipitation and the Generalized Logistic and GEV distributions for

726 streamflow, fit the transformed data reasonably well, with rejection rates of 5% or less.

727 However, in most cases these three-parameter distributions do not have a lower bound at

728 zero. This means that the lower tail of the fitted distribution may potentially go below zero,

729 which would result in a lower limit to the calculated SPI and SSI values (as observations can

730 never reach into this lower tail of the theoretical distribution). In contrast, the Tweedie

731 distribution is a flexible three-parameter distribution that has a lower bound at zero. It fits

732 both precipitation and streamflow data nearly as well as the best of the traditionally used

733 three-parameter distributions, with rejection rates of 2.4% and 3.9% for precipitation and

734 streamflow, respectively. The Tweedie distribution has only recently been applied to

735 precipitation data, and only for a few sites. As far as the authors are aware, it has never been

736 applied to streamflow data. The Tweedie probability density function can only be written in

737 closed form for a few special cases, which means that parameter estimation is time

738 consuming. However, recent advances in parameter estimation methods, and implementation

739 of these methods in the R package "tweedie" [*Dunn*, 2015], mean that the use of the Tweedie

740 distribution is now a viable option. The Tweedie distribution thus has significant potential for

741 use in future drought indicator applications in the UK, and has properties which make it well

742 suited for use in other environments. Particularly, its possibility of having mass at zero makes

743 it a proposition for drier climates. A further advantage is its encompassing of several other

744 distributions as special cases, which allows a greater flexibility in the possible shapes of the

745 distribution without having to make an a priori distributional choice. More generally, the

746 Tweedie could have utility for a range of other hydrological applications, suggesting further

747 research is needed to explore its full potential.

748   This study also investigated the effect of seasonality and catchment characteristics on the

749   goodness-of-fit of the statistical distributions. For both precipitation and streamflow data,

750   most distributions were found to have larger rejection rates in summer than in winter,

751   although for well-fitting distributions the differences were rather small. Similarly, catchment

752   characteristics generally had little impact on the suitability of a distribution.

753

763

764

References

Bachmair, S., K. Stahl, K. Collins, J. Hannaford, M. Acreman, M. Svoboda, C. Knutson, K. Helm Smith, N. Wall, B. Fuchs, N. D. Crossman, and I. C. Overton (2016), Drought indicators revisited: the need for a wider consideration of environment and society, *Wiley Interdisciplinary Reviews: Water 3*(4), 516-536, doi: 10.1002/wat2.1154.

Barker, L. J., J. Hannaford, A. Chiverton, and C. Svensson (2016), From meteorological to hydrological drought using standardised indicators, *Hydrol. Earth Syst. Sci., 20*, 2483-2505, doi: 10.5194/hess-20-2483-2016.

Bayliss, A. (1999), Flood Estimation Handbook vol. 5: Catchment descriptors, Institute of Hydrology, Wallingford, UK.

Bloomfield, J. P., and B. P. Marchant (2013), Analysis of groundwater drought building on the standardised precipitation index approach, *Hydrol. Earth Syst. Sci.*, *17*(12), 4769-4787, doi:10.5194/hess-17-4769-2013.

Bradford, R. B., and T. J. Marsh (2003), Defining a network of benchmark catchments for the UK, *P. I. Civil Eng. – Water & Mar. En., 156*(WM2), 109-116.

Chiverton, A., J. Hannaford, I. Holman, R. Corstanje, C. Prudhomme, J. Bloomfield, and T. M. Hess (2015), Which catchment characteristics control the temporal dependence structure of daily river flows? *Hydrol. Process.*, *29*(6), 1353-1369, doi: 10.1002/hyp.10252.

Collins, K., J. Hannaford, S. Haines, L. Stephens (2015), Drought: understanding and reducing vulnerability through monitoring and early warning systems. Report of the DrIVER workshop, 17 March 2015, Wallingford, UK.

Dunn, P. K. (2004), Occurrence and quantity of precipitation can be modelled simultaneously, *Int. J. Climatol., 24*(10), 1231-1239, doi: 10.1002/joc.1063.

Dunn, P. K. (2015), Package 'tweedie'. Description of the R programming language package 'tweedie', http://cran.r-project.org/web/packages/tweedie/tweedie.pdf (link accessed on 14 September 2016).

Dunn, P. K. and G. K. Smyth (2001), Tweedie family densities: methods of evaluation, Proc. 16th International Workshop on Statistical Modelling, Odense, Denmark, 2-6 July 2001, 8 pp.

793   Dunn, P. K. and G. K. Smyth (2005), Series evaluation of Tweedie exponential dispersion
794   model densities, *Stat. Comput., 15*(4), 267-280, doi: 10.1007/s11222-005-4070-y.

795   Dunn, P. K. and G. K. Smyth (2008), Evaluation of Tweedie exponential dispersion model
796   densities by Fourier inversion, *Stat. Comput., 18*(1), 73-86, doi: 10.1007/s11222-007-9039-6.

797   Farrell, P. J. and K. Rogers-Stewart (2006), Comprehensive study of tests for normality and
798   symmetry: extending the Spiegelhalter test. *J. Stat. Comput. Sim., 76*(9), 803-816, doi:
799   10.1080/10629360500109023.

800   Folland, C. K., J. Hannaford, J. P. Bloomfield, M. Kendon, C. Svensson, B. P. Marchant, J.
801   Prior, and E. Wallace (2015), Multi-annual droughts in the English Lowlands: a review of
802   their characteristics and climate drivers in the winter half-year, *Hydrol. Earth Syst. Sci.*,
803   *19*(5), 2353-2375, doi: 10.1080/10629360500109023.

804   Gosling, R. (2014), Assessing the impact of projected climate change on drought
805   vulnerability in Scotland, *Hydrol. Res.*, *45*(6), 806-816, doi: 10.2166/nh.2014.148.

806   Gudmundsson, L., and J. H. Stagge (2015), Package 'SCI'. Description of the R
807   programming language package 'SCI', http://cran.r-project.org/web/packages/SCI/SCI.pdf
808   (link accessed on 14 September 2016).

809   Guttman, N. B. (1999), Accepting the standardized precipitation index: a calculation
810   algorithm. *J. Am. Water Resour. As., 35*(2), 311-322, doi: 10.1111/j.1752-
811   1688.1999.tb03592.x.

812   Hannaford, J., B. Lloyd-Hughes, C. Keef, S. Parry, and C. Prudhomme (2011), Examining
813   the large-scale spatial coherence of European drought using regional indicators of
814   precipitation and streamflow deficit, *Hydrol. Process.*, *25*(7), 1146-1162, doi:
815   10.1002/hyp.7725.

816   Hasan, M. M. and P. K. Dunn (2011), Two Tweedie distributions that are near-optimal for
817   modelling monthly rainfall in Australia. *Int. J. Climatol., 31*, 1389-1397, doi:
818   10.1002/joc.2162.

819   Hayes, M., M. Svoboda, N. Wall, and M. Widhalm (2011), The Lincoln Declaration on
820   Drought Indices, *B Am Meteorol Soc*, *92*(4), 485-488, doi: 10.1175/2010BAMS3103.1.

821   Hosking, J. R. M. (2015), Package 'lmom'. Description of the R programming language

822   package 'lmom', https://cran.r-project.org/web/packages/lmom/lmom.pdf (link accessed on

823   14 September 2016.

824   Hosking, J. R. M., and J. R. Wallis (1997) Regional frequency analysis – An approach based

825   on L-moments, Cambridge University Press, Cambridge, UK.

826   Jørgensen, B. (1987), Exponential dispersion models, *J. R. Statist. Soc. B, 49*(2), 127-162.

827   Jørgensen, B., and M. C. Paes de Souza (1994), Fitting Tweedie's compound Poisson model

828   to insurance claims data, *Scand. Actuar. J., 1994*(1), 69-93, doi:

829   10.1080/03461238.1994.10413930.

830   Lloyd-Hughes, B. (2014), The impracticality of a universal drought definition, *Theor. Appl.*

831   *Climatol.*, *117*(3-4), 607-611, doi: 10.1007/s00704-013-1025-7.

832   Lloyd-Hughes, B. and M. A. Saunders (2002), A drought climatology for Europe, *Int. J.*

833   *Climatol., 22*(13), 1571-1592, doi: 10.1002/joc.846.

834   Lorance, P., L. Pawlowski, and V. M. Trenkel (2010), Standardizing blue ling landings per

835   unit effort from industry haul-by-haul data using generalized additive models, *ICES J. Mar.*

836   *Sci.*, *67*(8), 1650-1658, doi: 10.1093/icesjms/fsq048.

837   Marsh, T. J. and J. Hannaford (Eds.) (2008), UK Hydrometric Register. Hydrological data

838   UK series, Centre for Ecology & Hydrology, Wallingford, UK, 210 pp.

839   McKee, T. B., N. J. Doelsken and J. Kleist (1993), The relationship of drought frequency and

840   duration to time scales. Proc. Eighth Conference on Applied Climatology, 17-22 January

841   1993, Anaheim, California, 6 pp.

842   Núñez, J., D. Rivera, R. Oyarzún, and J. L. Arumí (2014), On the use of Standardized

843   Drought Indices under decadal climate variability: Critical assessment and drought policy

844   implications, *J. Hydrol.*, *517*, 458-470, doi: 10.1016/j.jhydrol.2014.05.038.

845   Razali, N. M. and Y. B. Wah (2011), Power comparisons of Shapiro-Wilk, Komogorov-

846   Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statististical Modeling and*

847   *Analysis, 2*(1), 21-33.

848   Shukla, S., and A. W. Wood (2008), Use of a standardized runoff index for characterizing

849   hydrologic drought, *Geophys. Res. Lett.*, *35*(2), L02405, doi: 10.1029/2007GL032487.

850 Sienz, F., O. Bothe, and K. Fraedrich (2012), Monitoring and quantifying future climate

851 projections of dryness and wetness extremes: SPI bias, *Hydrol. Earth Syst. Sci.*, *16*(7), 2143-

852 2157, doi:10.5194/hess-16-2143-2012.

853 Soláková, T., C. De Michele, and R. Vezzoli (2014), Comparison between Parametric and

854 Nonparametric Approaches for the Calculation of Two Drought Indices: SPI and SSI, *J.*

855 *Hydrol. Eng.*, *19*(9), 04014010, doi: 10.1061/(ASCE)HE.1943-5584.0000942.

856 Stagge, J. H., L. M. Tallaksen, L. Gudmundsson, A. F. van Loon, and K. Stahl (2015),

857 Candidate distributions for climatological drought indices (SPI and SPEI), *Int. J. Climatol.*,

858 *35*(13), 4027-4040, doi: 10.1002/joc.4267.

859 Stagge, J. H., L. M. Tallaksen, L. Gudmundsson, A. F. van Loon, and K. Stahl (2016),

860 Response to comment on 'Candidate distributions for climatological drought indices (SPI and

861 SPEI)', *Int. J. Climatol.*, *36*(4), 2132-2138, doi: 10.1002/joc.4564.

862 Staudinger, M., K. Stahl, and J. Seibert (2014), A drought index accounting for snow, *Water*

863 *Resour. Res.*, *50*(10), 7861-7872, doi: 10.1002/2013WR015143.

864 Tanguy, M., J. Hannaford, L. Barker, C. Svensson, F. Kral, and M Fry (2015) Gridded

865 Standardised Precipitation-Evapotranspiration Index (SPEI) using generalised logistic

866 distribution with standard period 1961-2010 for Great Britain [SPEIgenlog61-10]

867 doi:10.5285/d201a2af-568e-4195-bf02-961fb6954c72 – Supporting Information for dataset,

868 NERC Environmental Information Data Centre, UK, 8 pp.

869 Tweedie, M. C. K. (1981), An index which distinguishes between some important

870 exponential families. In Statistics: Applications and New Directions, Proceedings of the

871 Indian Statistical Institute Golden Jubilee International Conference, edited by Ghosh, J. K.

872 and Roy, J., pp. 579-604, Indian Statistical Institute, Calcutta, India.

873 Vicente-Serrano, S. M., and S. Beguería (2015), Comment on 'Candidate distributions for

874 climatological drought indices (SPI and SPEI)' by James H. Stagge *et al.*, *Int. J. Climatol.,*

875 *36*(4), 2120-2131, doi: 10.1002/joc.4474.

876 Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2010), A Multiscalar Drought

877 Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration

878 Index, *J. Climate*, *23*(7), 1696-1718, doi: 10.1175/2009JCLI2909.1.

879   Vicente-Serrano, S. M., J. I. López-Moreno, S. Beguería, J. Lorenzo-Lacruz, C. Azorin-

880   Molina, and E. Morán-Tejeda (2012), Accurate Computation of a Streamflow Drought Index,

881   *J. Hydrol. Eng.*, *17*(2), 318-332, doi: 10.1061/(ASCE)HE.1943-5584.0000433.

882   Wilby, R. L., C. Prudhomme, S. Parry, and K. G. L. Muchan (2015), Persistence of

883   hydrometeorological droughts in the United Kingdom: A regional analysis of multi-season

884   rainfall and river flow anomalies, *J. Extreme Events*, *2*(2), 1550006 (27pp), doi:

885   10.1142/S2345737615500062.

886   Wilhite, D., and M. Glantz (1985), Understanding: the drought phenomenon: the role of

887   definitions, *Water Int*.,*10*(3), 111-120, doi:10.1080/02508068508686328.

888   Wu, H., M. J. Hayes, D. A. Wilhite, and M. D. Svoboda (2005), The effect of the length of

889   record on the standardized precipitation index calculation, *Int. J. Climatol*, *25*(4), 505-520,

890   doi: 10.1002/joc.1142.

891   Wu, H., M. D. Svoboda, M. J. Hayes, D. A. Wilhite, and F. J. Wen (2007), Appropriate

892   application of the Standardized Precipitation Index in and locations and dry seasons, *Int. J.

893   Climatol.*, *27*(1), 65-79, doi: 10.1002/joc.1371.

894

895

Appendix A. The Tweedie family of distributions

The concept of the Tweedie family of distributions was comprehensively outlined by *Tweedie* [1984], and further explored by *Jørgensen* [1987]. The Tweedie family of distributions is a special class within the so called exponential dispersion models (EDMs). A random variable *Y* that follows an EDM has a probability density function on the form

$$f(y;\mu,\phi) = a(y,\phi)\exp\left[\frac{1}{\phi}\{y\theta - \kappa(\theta)\}\right],$$

where $\mu = E[Y] = \kappa'(\theta)$ is the mean, $\phi > 0$ is the dispersion parameter, θ is the canonical parameter and $\kappa(\theta)$ is the cumulant function (the latter two are both known). The function $a(y,\phi)$ cannot generally be written in closed form, apart from for some special cases. The variance is given by $\mathrm{var}[Y] = \phi V(\mu)$ where $V(\mu) = \kappa''(\theta)$ is the variance function viewed as a function of the mean μ. The Tweedie family of distributions is a class of EDMs characterized by the variance function $V(\mu)$ having a power relationship to the mean, μ, as $V(\mu) = \mu^\xi$ for $\xi \notin (0,1)$, where ξ specifies the particular distribution. Special cases include the Normal ($\xi = 0$), Poisson ($\xi = 1$), Gamma ($\xi = 2$), and Inverse Gaussian ($\xi = 3$) distributions [e.g. *Dunn and Smyth*, 2001; *Hasan and Dunn*, 2011]. Apart from for these special cases, numerical methods need to be used for estimating the distribution parameters [e.g. *Dunn and Smyth*, 2001; 2008]. Although time consuming compared with parameter estimation for distributions with explicit analytic forms, the implementation of such methods in the R package 'tweedie' [*Dunn*, 2015] means that application of the Tweedie distribution is now a more practicable option.

When the Tweedie parameter ξ has a value between 1 and 2, the distribution describes series that contain exact zeroes as well as positive continuous data. *Dunn* [2004] refers to this particular case as a Poisson-gamma distribution, and explains the underlying assumptions when applying it to precipitation data as follows. If the number of precipitation events occurring within a time period is assumed to have a Poisson distribution, and the magnitude of each event is Gamma distributed, then the total precipitation amount within the time period can be found as the Poisson sum of the Gamma distributed random variables. This distribution has also been referred to as a compound Poisson distribution by, for example, *Jørgensen and Paes de Souza* [1994] who used it to model total insurance claims costs per

925   insured unit (assuming the number of individual claims within the unit to be Poisson

926   distributed and the cost for each individual claim to be Gamma distributed).

927   The EDMs are a generalization of (linear) exponential families, and describe the error

928   distribution of generalized linear models (GLMs) [e.g. *Jørgensen*, 1987]. *Hasan and Dunn*

929   [2011] note that a framework is therefore already in place for fitting GLMs based on the

930   Tweedie distributions, and for diagnostic testing. Further, it is straight-forward to incorporate

931   covariates into the modelling procedure. This has been done by, for example, *Lorance et al.*

932   [2010] for estimating fish landings per unit effort using generalized additive models with

933   explanatory variables such as depth, area and year.

934

935

Appendix B. Fitting statistical distribution functions

A range of two- and three-parameter probability distributions was investigated, as well as one four-parameter distribution: the Kappa distribution. The three-parameter distributions included the Generalized Logistic, Pearson type 3, Generalized Extreme Value (GEV) and Tweedie, as well as the Kappa distribution with a lower bound at zero imposed (which reduces it to a three-parameter distribution). The two-parameter distributions included the Gamma, Lognormal, Normal, Gumbel and Weibull distributions, as well as the Generalized Logistic distribution with a lower bound at zero (which reduces it to a two-parameter distribution). The definitions of the probability density functions are given in *Stagge et al.* [2015], except for the Tweedie which is described by *Dunn and Smyth* [2005] and the Kappa which is described in *Hosking and Wallis* [1997]. The constraints of a lower bound at zero were implemented based on the equations for the lower bounds given in *Hosking and Wallis* [1997]. It can be noted that the distribution that is referred to as the Generalized Logistic in the present paper and in *Stagge et al.* [2015], follow the definition in *Hosking and Wallis* [1997]. Others, for example *Vicente-Serrano et al.* [2010] and *Vicente-Serrano and Beguería* [2015], refer to this same distribution as the Log-logistic distribution.

Computationally, the aggregation of the data into durations longer than a month and the fitting of the distributions to the aggregated series were done using the function fitSCI in the R package 'SCI' (which calls other packages to do the actual fits) [*Gudmundsson and Stagge*, 2015]. All of the distributions except the Tweedie, the Kappa and the distributions bounded below at zero are available within the 'SCI' package. In order to incorporate these latter distributions into the 'SCI' framework, new start functions for parameter estimates were written (start.fun in the fitSCI function). For all distributions except the Tweedie, the starting values for the maximum likelihood estimation of the parameter values were based on L-moments. The initial, complete, set of L-moment-based parameter estimates were obtained according to *Hosking and Wallis* [1997], using the samlmu, pelglo and pelkap functions in the R package 'lmom' [*Hosking*, 2015]. For the constrained distributions, these initial parameter estimates were then amended to make the distributions conform to having a lower bound at zero, based on the equations for the lower bounds given in *Hosking and Wallis* [1997]. For the Generalized Logistic distribution, the location and scale parameter estimates were kept, and the shape parameter, k, amended. Similarly, for the Kappa distribution, all L-moment-based parameter estimates except the parameter k were kept, and k was amended.

968   For the Tweedie distribution, the start function obtained the parameter estimates by calling

969   the function tweedie.profile in the R package 'tweedie' [*Dunn*, 2015]. The Tweedie

970   parameters μ (the mean), ξ (the power) and ϕ (the dispersion) are estimated using maximum

971   likelihood as described by *Dunn* [2015] and *Dunn and Smyth* [2005], using their suggested

972   order of preference regarding estimation method. When ξ is close to one, the probability

973   density function becomes multimodal, reflecting that at this limit it is a Poisson distribution

974   which is discrete [*Dunn and Smyth,* 2005]. The value of the parameter ξ was therefore

975   constrained. *Dunn* [2015] recommends a minimum ξ of 1.2 for data containing zeroes and 1.5

976   for data without zeroes. The higher limit results in fewer distributions being fitted with mass

977   at zero ($F(0)>0$). However, there is a trade-off with the fit of the distributions, as particularly

978   for longer aggregation durations the distributions become more symmetrical and the optimum

979   value of ξ decreases towards 1. In the present study $\xi \geq 1.2$ has been used as we found the

980   values of $F(0)$ in Table 3 acceptable.

981   If possible, the parameter values for the distributions available in 'SCI' are estimated using

982   maximum likelihood. However, if the maximum likelihood method fails to converge, an

983   attempt is first made to estimate the parameter values using L-moments, and failing that too,

984   the method of moments is used. For the bounded distributions and the four-parameter Kappa

985   distribution, L-moment estimators were used if the maximum likelihood procedure did not

986   converge. For most distributions, the maximum likelihood procedure failed to converge for

987   (much) less than one percent of cases. However, for the GEV and the four-parameter Kappa

988   distribution the failure rates were 2.6% and 3.4% (GEV) and 23% and 40% (Kappa) for

989   precipitation and river flow data, respectively.

990   Once the distributions had been fitted, the standardized indices were calculated using the

991   function transformSCI in the R package 'SCI'. The function ptweedie in the 'tweedie'

992   package was amended to cope with i) the SCI function transformSCI passing the vector of

993   quantiles to ptweedie as mode list rather than numeric, and ii) the vector of quantiles

994   containing missing values.

995

996

997   **Tables**

998

999   **Table 1.** Statistical probability distributions investigated. Bold italic font denote distributions with a lower bound at zero. Arrows indicate where
1000   a distribution with fewer parameters is a special case or variant of a more general distribution.

1001

4 parameters (very flexible)	3 parameters (flexible)		2 parameters (less flexible)
	Generalized Logistic	→	*Generalised Logistic (lower bound imposed)*
	Pearson type 3	→	*Gamma*
	Generalized Extreme Value	→ {	*Weibull* Gumbel
	Tweedie	(→)*	Normal
			Lognormal
Kappa4	→	*Kappa3 (lower bound imposed)*	

1002   * The way the Tweedie distribution is constrained in the present study, it does not include the Normal distribution as an option.

1003

1004

1005 **Table 2.** Percentage of occurrences for which the normal distribution is rejected at the 5% significance level by the Shapiro-Wilk test for SPI
1006 and SSI of different aggregation durations, and for all durations together. Results are lumped together for the 121 catchments and the 12
1007 calendar end months. Distributions in bold italic font have a lower bound at zero.

Distribution	All SPI	SPI 1	SPI 3	SPI 6	SPI 12	All SSI	SSI 1	SSI 3	SSI 6	SSI 12
Gamma	*9.9*	*17.5*	*6.5*	*7.0*	*8.6*	*19.2*	*29.5*	*18.9*	*12.3*	*16.0*
Generalized Logistic	1.8	3.7	1.4	1.6	0.6	3.8	9.0	3.7	1.6	0.9
Pearson type 3	1.2	2.1	0.6	0.9	1.2	10.5	25.9	11.6	3.5	1.0
GEV	3.5	1.2	1.2	3.9	7.6	4.7	6.6	2.8	3.0	6.4
Tweedie	*2.4*	*2.1*	*0.8*	*2.0*	*4.8*	*3.9*	*5.2*	*2.5*	*2.5*	*5.4*
Kappa3	*1.2*	*1.7*	*0.9*	*1.5*	*0.6*	*2.2*	*5.5*	*1.7*	*1.0*	*0.8*
Kappa4	0.7	1.2	0.4	0.8	0.5	1.3	3.1	1.3	0.3	0.3

1008

1009

1010

Table 3. Counts of $F(0)$ in different magnitude intervals. If $F(0)$ is large, then a large part of the fitted theoretical distribution occurs for data values at or below zero. Results are lumped for the 121 catchments, the four aggregation durations, and the 12 calendar end months.

| Magnitude of $F(0)$	Lower bound of SPI (or SSI) corresponding to $F(0)$	Precipitation					Streamflow				
		Generalized Logistic	Pearson type 3	GEV	Tweedie	Kappa4	Generalized Logistic	Pearson type 3	GEV	Tweedie	Kappa4
$F(0) \leq 0.01$	SPI $\leq$ -2.33	5358	5694	5557	5798	5746	5373	5653	5458	5791	5631
$0.01 < F(0) \leq 0.02$	-2.33 < SPI $\leq$ -2.05	322	100	216	10	40	255	106	248	17	98
$0.02 < F(0) \leq 0.03$	-2.05 < SPI $\leq$ -1.88	104	13	32	0	15	137	14	70	0	53
$0.03 < F(0) \leq 0.04$	-1.88 < SPI $\leq$ -1.75	22	1	3	0	6	31	0	20	0	20
$0.04 < F(0) \leq 0.05$	-1.75 < SPI $\leq$ -1.64	1	0	0	0	0	6	5	4	0	3
$0.05 < F(0) \leq 0.06$	-1.64 < SPI $\leq$ -1.55	1	0	0	0	0	1	2	2	0	3
$F(0) > 0.06$	SPI > -1.55	0	0	0	0	0	5	28	6	0	0

1016 **Table 4.** Correlation between seasonal average p-value from the Shapiro-Wilk goodness-of-fit test and different catchment descriptors, for 121

1017 catchments. Correlations significant at the 5% level are shown in bold font.

	Catchment descriptor	1-month streamflow						1-month precipitation				
		Gamma	Generalized Logistic	GEV	Tweedie	Kappa3	Kappa4	Gamma	Pearson type 3	Tweedie	Kappa3	Kappa4
Winter	BFIHOST	**-0.35**	0.11	-0.10	-0.05	-0.05	-0.04					
	SMDBAR*	-0.26	**-0.36**	**-0.56**	-0.34	**-0.37**	-0.24					
	SAAR	**0.44**	**0.27**	**0.45**	0.15	**0.34**	**0.35**	0.03	0.17	0.03	0.10	0.15
	AREA	**0.20**	0.01	0.13	0.14	-0.02	0.12	-0.05	-0.03	-0.03	-0.09	0.14
Summer	BFIHOST	-0.05	**0.46**	**0.23**	-0.05	0.17	0.11					
	SMDBAR*	-0.35	0.05	-0.14	-0.32	-0.10	**-0.43**					
	SAAR	**0.57**	**-0.31**	0.00	**0.34**	0.06	**0.34**	**-0.23**	**0.18**	-0.11	**0.33**	0.15
	AREA	0.07	-0.05	0.04	0.10	-0.14	0.12	0.09	**0.19**	0.17	0.12	0.03

1018 * SMDBAR is only available for 31 catchments.

Figures with captions
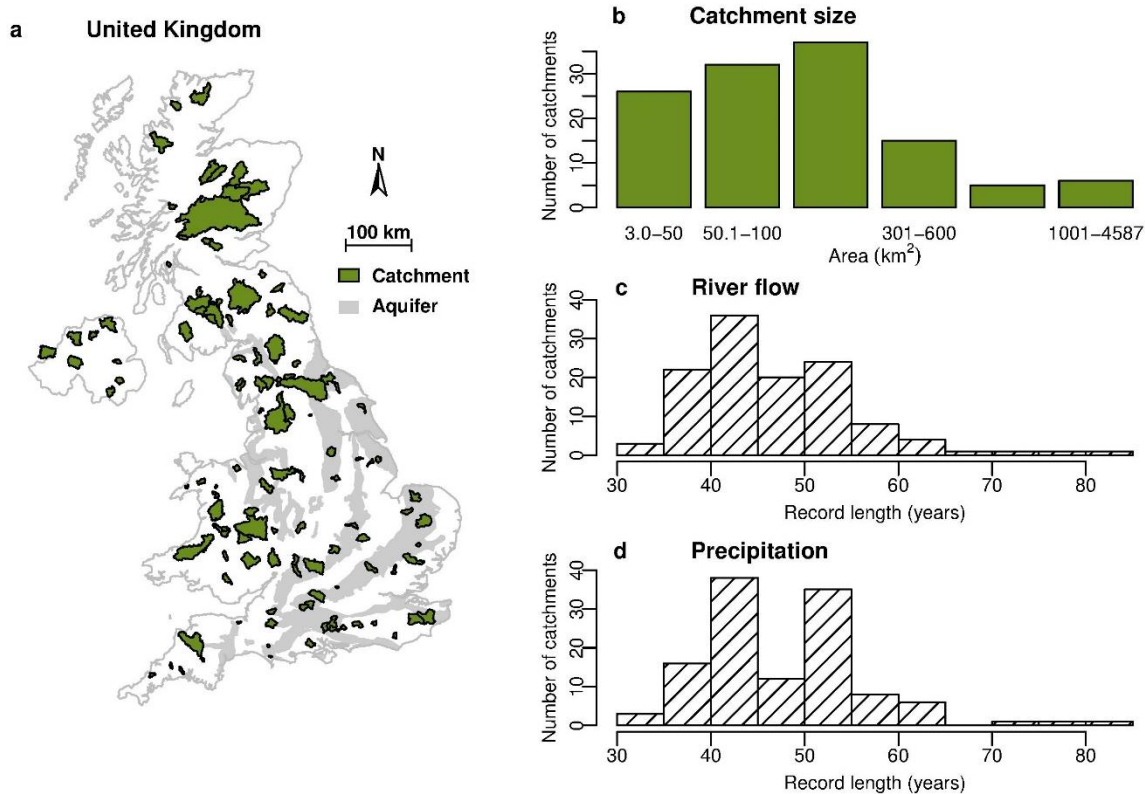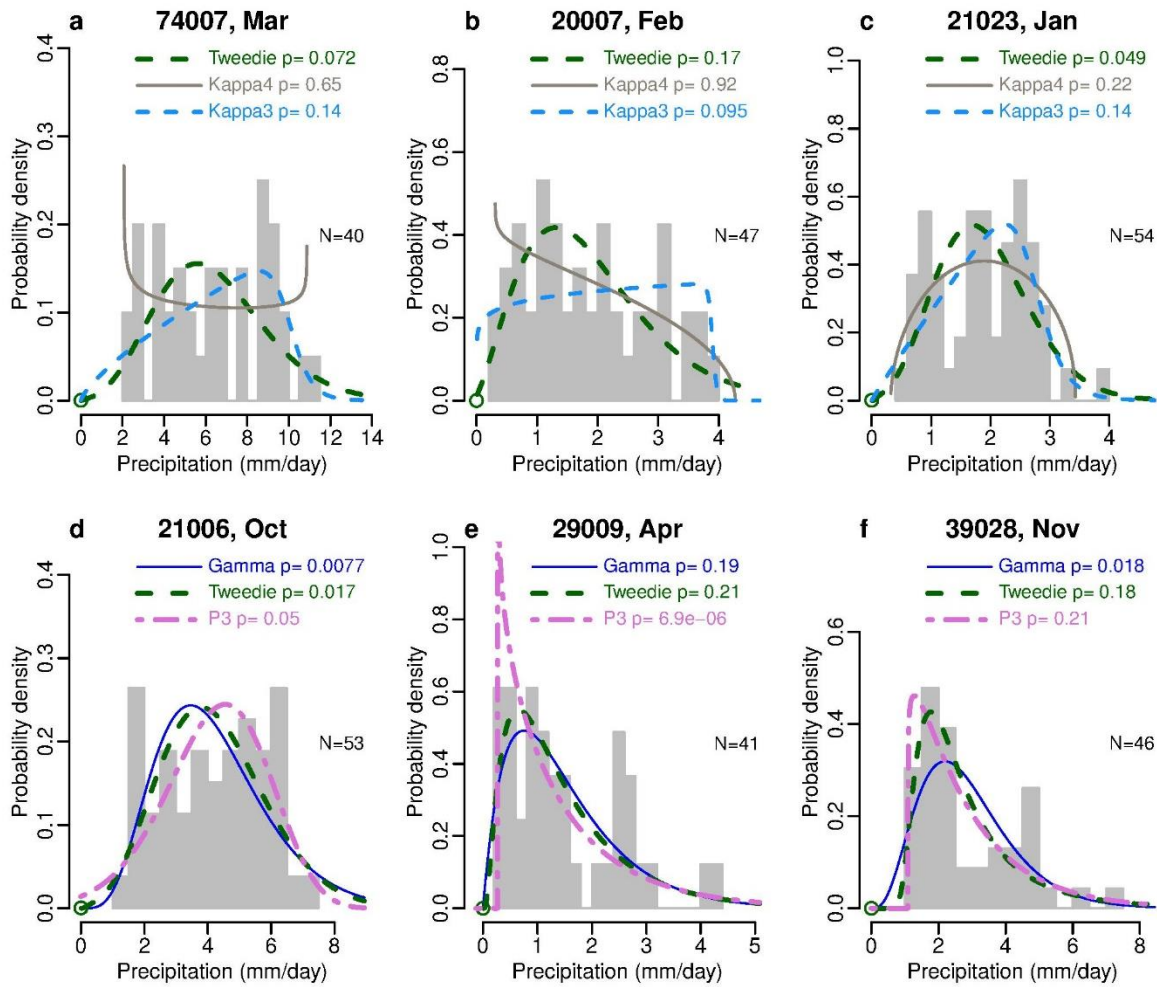
Figure 1. Location (a) and size (b) of the 121 study catchments in the UK. Record lengths

excluding missing data for streamflow (c) and precipitation (d).

Figure 2. Examples of histograms of observed monthly mean precipitation and fitted probability density functions for a) Esk at Cropple How (hydrometric number 74007), b) Gifford Water at Lennoxlove (20007), c) Leet Water at Coldstream (21023), d) Tweed at Boleside (21006), e) Ancholme at Toft Newton (29009), and f) Dun at Hungerford (39028). The green circle shows the mass at zero for the fitted Tweedie distribution, which for convenience is shown on the same axis as the probability density function although it is a dimensionless probability between 0 and 1.
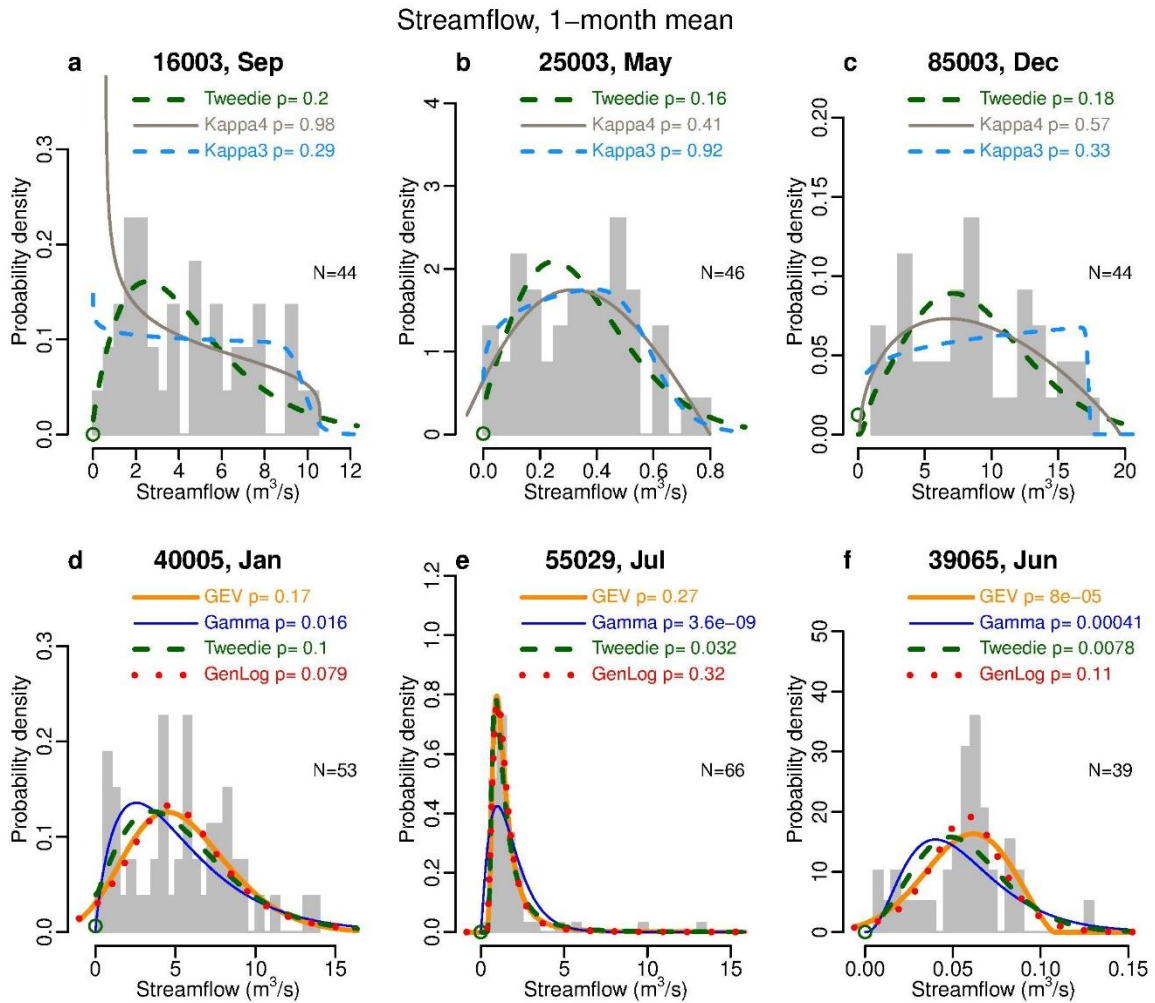
Figure 3. Examples of histograms of observed monthly mean streamflows and fitted probability density functions for a) Ruchill Water at Cultybraggan (hydrometric number 16003), b) Trout Beck at Moor House (25003), c) Falloch at Glen Falloch (85003), d) Beult at Stile Bridge (40005), b) Monnow at Grosmont (55029) and c) Ewelme Brook at Ewelme (39065). The green circle shows the mass at zero for the fitted Tweedie distribution, which for convenience is shown on the same axis as the probability density function although it is a dimensionless probability between 0 and 1.
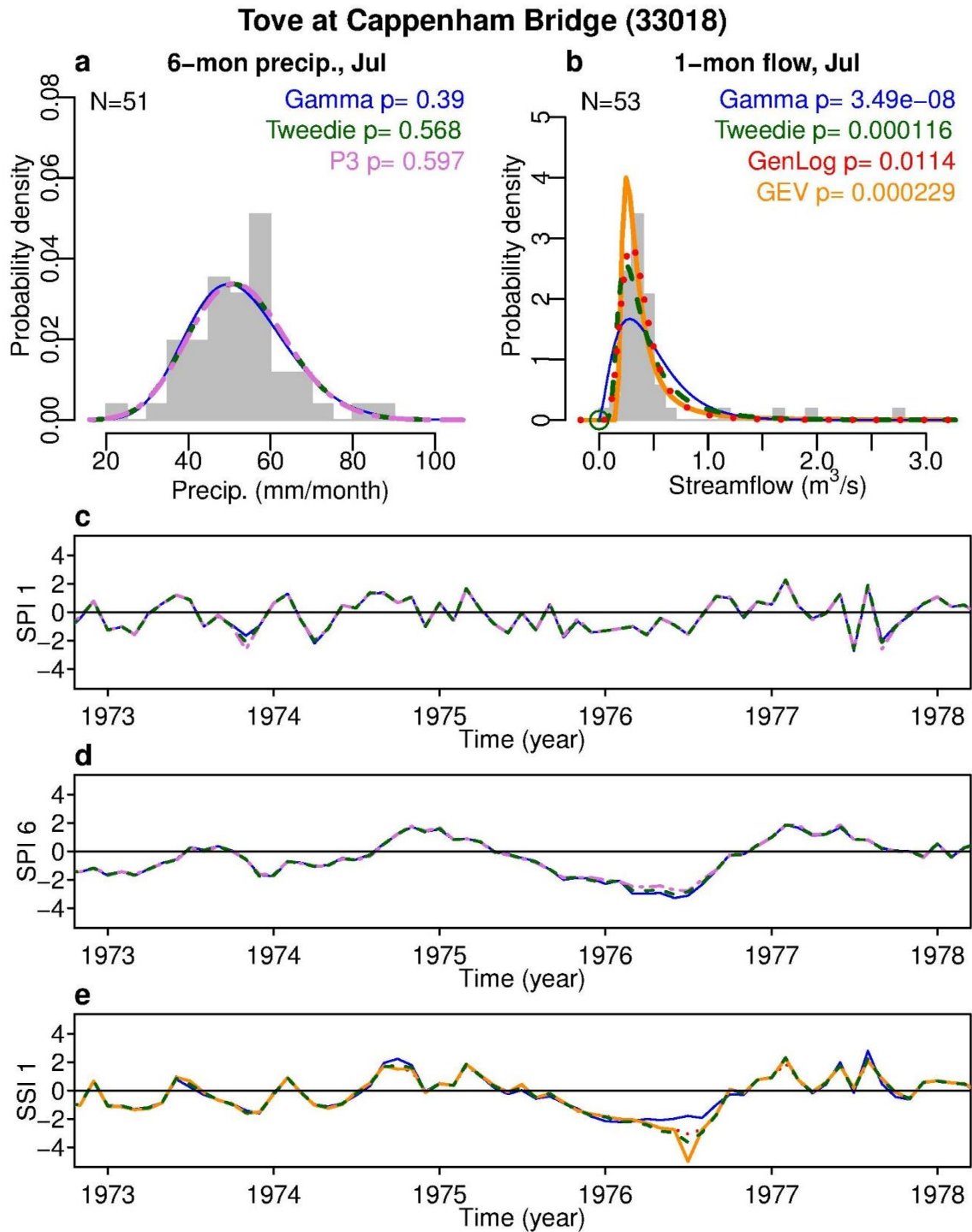
P−values ≥ 0.05 only
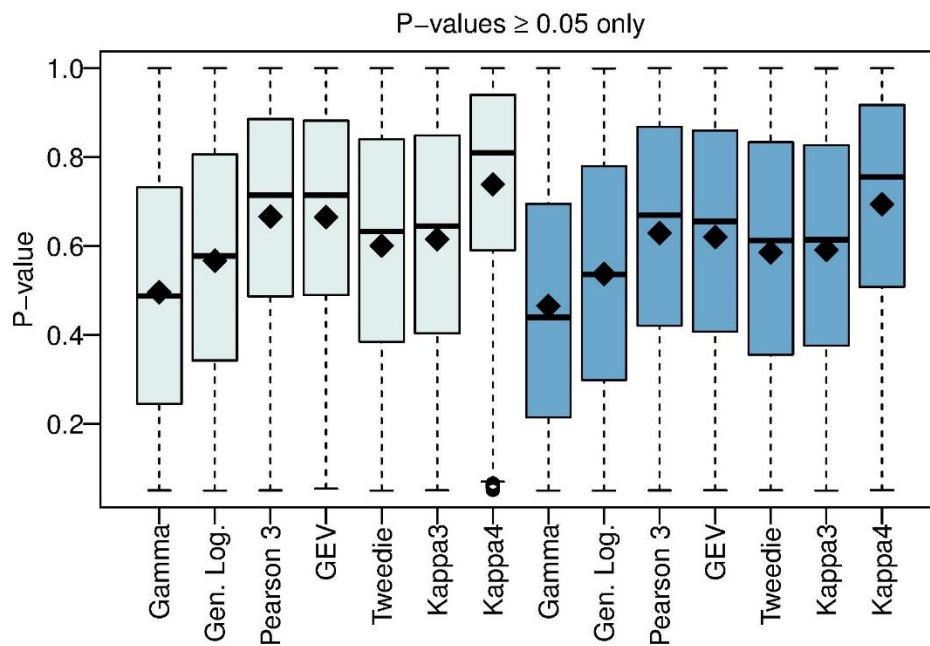
1045

1046   **Figure 4.** P-values for distributions accepted by the Shapiro-Wilk tests of normality for the

1047   SPI (light grey boxes) and SSI (blue boxes), for all 121 catchments, four aggregation

1048   durations, and 12 calendar end months lumped together. The box shows the inter-quartile

1049   range, the whiskers the maximum and minimum values, the thick line the median and the

1050   black diamond the mean of the p-values.

1051

Tove at Cappenham Bridge (33018)

a 6-mon precip., Jul

N=51

Gamma p= 0.39
Tweedie p= 0.568
P3 p= 0.597

b 1-mon flow, Jul

N=53

Gamma p= 3.49e-08
Tweedie p= 0.000116
GenLog p= 0.0114
GEV p= 0.000229

c

d

e

1052

Figure 5. Effect of distribution choice on the calculated SPI and SSI. Example histograms and fitted distributions (see Figures 2 and 3 for line types) for a) 6-month mean precipitation, and b) 1-month mean streamflow for the Tove at Cappenham Bridge (hydrometric number 33018), and time series for the c) SPI 1, d) SPI 6 and e) SSI 1 drought indicators.

1053
1054
1055
1056

1057
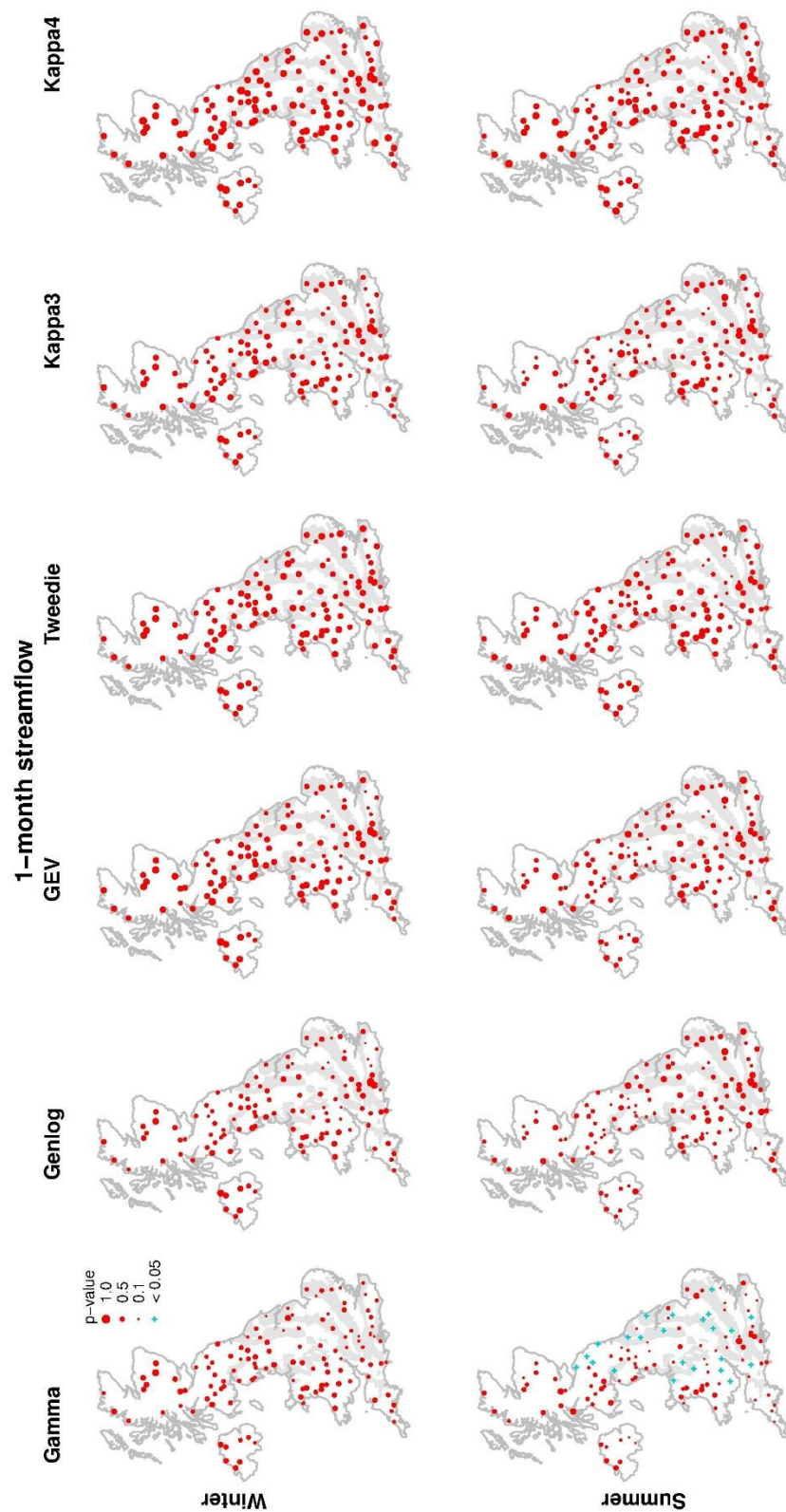
Figure 6. Seasonal average p-values for the Shapiro-Wilk goodness-of-fit test for SSI 1 at individual catchments, for a selection of distributions.