Journal of Hydrology 534 (2016) 659–668

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A two-component generalized extreme value distribution for precipitation frequency analysis

Zuzana Rulfová [a,b,c,*], Adri Buishand [d], Martin Roth [d], Jan Kyselý [a,e]

[a] Institute of Atmospheric Physics CAS, Boční II 1401, 141 31 Prague 4, Czech Republic
[b] Dept. of Applied Mathematics, Technical University of Liberec, Univerzitní náměstí 1410/1, 461 17 Liberec, Czech Republic
[c] Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 3, 121 16 Prague 2, Czech Republic
[d] Royal Netherlands Meteorological Institute (KNMI), Postbus 201, 3730 AE De Bilt, Netherlands
[e] Faculty of Environmental Sciences, Czech University of Life Sciences, Kamýcká 1176, 165 21 Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form 15 December 2015
Accepted 16 January 2016
Available online 23 January 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Felix Frances, Associate Editor

Keywords:
Precipitation extremes
Two-component extreme value distribution
Regional frequency analysis
Convective precipitation
Stratiform precipitation
Central Europe

SUMMARY

A two-component generalized extreme value (TCGEV) distribution is introduced based on the assumption that the annual maxima for convective and stratiform precipitation follow two separate generalized extreme value (GEV) distributions. The regional TCGEV model is used to analyze 6-h precipitation data for 11 stations in the Czech Republic over 1982–2010 subdivided into predominantly convective and stratiform precipitation. For each type of precipitation, the shape parameter and the ratio of the scale parameter and the location parameter of the underlying GEV distributions are assumed to be constant over the region. The validity of this homogeneity assumption is explored with a bootstrap procedure and the goodness-of-fit is tested with the Anderson–Darling statistic both for each individual station and for all stations simultaneously. The return levels from the regional TCGEV distribution are compared with those obtained with the common method of fitting a regional GEV distribution to the overall annual maxima, ignoring their convective or stratiform origin. The differences are generally small, but they increase with return period and are larger at lowland stations where the proportion of convective precipitation extremes is greater. High return levels based on a GEV fit to the overall annual maxima for these stations tend to be smaller than those for the convective component owing to the heavier upper tail of the distribution of convective extremes. Results from the TCGEV distribution are consistent, i.e., the estimated return levels of the overall annual maxima cannot be smaller than those for the convective and stratiform components obtained from the GEV distribution.

1. Introduction

Characteristics of precipitation extremes are important in many practical applications, including hydrological modeling, design of hydraulic structures, urban planning, etc. Numerous studies have examined the distributions of precipitation extremes in observed data and climate model simulations, using methods of local or regional frequency analysis of different complexity (e.g., Stewart et al., 1999; Hanel et al., 2009; Svensson and Jones, 2010). One of the main assumptions of extreme value theory is that the maxima belong to the same population (see Coles, 2001). However, this assumption is not fulfilled for extreme precipitation in mid-latitudes, which is caused by two different physical mechanisms. Precipitation extremes arise from convective processes at small spatial scales (convective precipitation) or from cloud belts associated with cyclones and atmospheric fronts at larger scales (stratiform precipitation).

Where observations are generated by two independent processes, the distribution function of the overall extremes is the product of the distribution functions $F_1(x)$ and $F_2(x)$ of the extremes from the separate processes (e.g., Gumbel, 1958; Canfield et al., 1980; Waylen and Woo, 1982; Tabony, 1983):

$$F(x) = F_1(x) \cdot F_2(x) \quad (1)$$

Gumbel (1958) suggested the use of such a two-component distribution to analyze floods arising from snow melt in spring and precipitation in autumn. Waylen and Woo (1982) fitted Gumbel distributions to the annual maxima of snowmelt and precipitation generated floods, and concluded that the resulting two-component Gumbel distribution provided a good fit to the overall annual maximum floods. The two-component Gumbel distribution has also

* Corresponding author at: Institute of Atmospheric Physics CAS, Boční II 1401, 141 31 Prague 4, Czech Republic.

E-mail addresses: rulfova@ufa.cas.cz (Z. Rulfová), adri.buishand@knmi.nl (A. Buishand), roth@knmi.nl (M. Roth), kysely@ufa.cas.cz (J. Kyselý).

http://dx.doi.org/10.1016/j.jhydrol.2016.01.032
0022-1694/

been used for flood frequency modeling by Rossi et al. (1984), Strupczewski et al. (2012) and Kochanek et al. (2012), and for describing the distribution of extreme wind speeds (e.g., Gomes and Vickery, 1978; van den Brink et al., 2004; Escalante-Sandoval, 2008) and precipitation extremes for different durations (e.g., Caporali et al., 2008). In a number of these applications, however, the four parameters of the distribution were estimated jointly by fitting the distribution to the overall annual maxima, because an a priori subdivision of the observations according to the two generating mechanisms was not feasible. This leads in general to a larger standard error of quantile estimates (Strupczewski et al., 2012).

In this study, we analyze annual maxima of 6-h precipitation amounts in the Czech Republic (Central Europe) over the period 1982–2010 using a two-component generalized extreme value (TCGEV) distribution, which consists of two generalized extreme value (GEV) distributions (e.g., Coles, 2001). As sub-daily precipitation extremes have usually a longer upper tail than the Gumbel distribution (e.g., Alila, 1999; Overeem et al., 2008), use of the GEV distribution as basis for a two-component model for extreme precipitation is more appropriate. For parameter estimation of the TCGEV distribution, we take advantage of a subdivision of observed precipitation into predominantly convective and stratiform type based on surface weather observations from a previous study (Rulfová and Kyselý, 2013). The return levels from the TCGEV distribution are compared with those obtained by the common practice of fitting a GEV distribution to the overall annual maxima. We apply a regional frequency analysis, which assumes that the most uncertain distribution parameters are constant over the study area. The advantage of a regional frequency analysis is that sampling variations in the estimates of model parameters and high return levels can be substantially reduced compared to a single-site analysis (e.g., Lettenmaier et al., 1987; Cunnane, 1988; Stedinger et al., 1993). The homogeneity assumption is tested with a bootstrap procedure that takes spatial dependence into account. In addition, goodness-of-fit of the regional model is tested for each individual station and for all stations simultaneously, using the Anderson–Darling statistic.

The paper is structured as follows: After a short description of the data and precipitation patterns in the Czech Republic in Section 2, the extreme value models and parameter estimation are discussed in Section 3. Spatial homogeneity of distribution parameters and testing goodness-of-fit are dealt with in Section 4. The return levels from the GEV and TCGEV distributions are compared in Section 5. Section 6 presents a discussion and conclusions.

2. Data

2.1. Input dataset

The observed 6-h precipitation data originate from SYNOP reports at 11 stations (operated by the Czech Hydrometeorological Institute) over 1982–2010. The stations are approximately evenly distributed over the study area (Fig. 1) and cover different climatological regions within the Czech Republic, from lowlands to mountains (the altitudes of the stations range from 241 m to 1322 m a.s.l., Table 1). The data were quality checked and compared with other available datasets to identify gross errors and suspicious readings (Rulfová and Kyselý, 2013). Nine out of the 11 stations have less than 0.1% of missing 6-h precipitation data. Exceptions are the stations 11698 (Kuchařovice), with 4 months of missing values (January–April 1989), and 11406 (Cheb), with 3 months of missing values (October–December 1993). As 6-h precipitation extremes mainly occur in summer, it is assumed that the selected maxima in these incomplete years are the true annual maxima.

The time series of 6-h convective and stratiform precipitation amounts were obtained using the algorithm proposed and evaluated in detail in Rulfová and Kyselý (2013). It subdivides 6-h precipitation amounts into convective and stratiform type primarily based on hourly weather state data. The 6-h precipitation amounts are classified as convective if there are weather states corresponding to convective precipitation and no weather states corresponding with stratiform precipitation in the 6-h interval or if there are weather states corresponding to *heavy* convective precipitation in combination with weather states corresponding to *light* stratiform precipitation only. A similar criterion is used to classify the 6-h precipitation as stratiform. In cases when both convective and stratiform precipitation occurred within the 6-h interval, and there is no indication that the contribution from one of the types is negligible (or the data on weather state is missing for the 6-h interval), a secondary criterion based on cloud type is applied. Convective precipitation is associated with Cumulonimbus and Cumulus Congestus clouds while stratiform precipitation is mainly associated with Nimbostratus, but also with Stratocumulus, Stratus and Altostratus clouds. A small percentage of precipitation amounts (around 5%) remains unresolved even after application of the secondary criterion, so they are classified as mixed (Rulfová and Kyselý, 2013). Extreme 6-h precipitation classified as mixed comprises no more than one annual maximum at each station over the 29-year period (Table 2). These mixed maxima are taken into account in the analysis of the overall annual maxima but they are omitted from the analysis of the convective and stratiform components.

Eq. (1) assumes two independent processes; therefore we tested the independence of the 6-h annual maxima of convective and stratiform precipitation using Kendall's tau rank correlation coefficient (e.g., Wilks, 2006). No significant correlation at the 5% level was found.

2.2. Basic precipitation climatology of the study area

The Czech Republic is characterized by large spatial and temporal variability in precipitation. The annual cycle of mean monthly precipitation has a single maximum in June and July and minimum in January and February (Tolasz et al., 2007). The proportion of convective precipitation is small in winter and autumn (below 15%), becomes larger in spring (around 30%) and is about 50% in summer at most stations (Rulfová and Kyselý, 2013). While the number of wet days is greater in the cold half of the year, precipitation intensity is higher in the warm half of the year (Tolasz et al., 2007). Spatial variability of seasonal mean precipitation over the study area is linked to the large-scale atmospheric circulation and is modified locally by such factors as orography, wind exposure, precipitation shadow, and orientation of mountain range relative to the prevailing wind direction.

Sub-daily precipitation extremes occur in relation to thunderstorms (convective precipitation) or cloud belts associated with cyclones that usually originate in the Mediterranean area (stratiform precipitation, e.g., Štekl et al., 2001). Heavy precipitation at lowland stations is mostly of convective origin, while at high altitudes and at stations influenced by nearby mountains, intense stratiform precipitation becomes more important (Rulfová and Kyselý, 2013). Box plots of the 6-h annual maxima for convective and stratiform precipitation and the overall 6-h annual maxima are depicted in Fig. 2. The distributions are usually positively skewed and the distribution of the overall 6-h maxima is similar to that for convective precipitation especially at stations where more than 60% of the extremes come from convective precipitation (Table 2). Mountain station 11787 (Lysá hora) is the only station with predominantly stratiform extremes (69% of the annual extremes is of stratiform origin for that station).