Journal of Hydrology 553 (2017) 49–58

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

A systematic approach to selecting the best probability models for annual maximum rainfalls – A case study using data in Ontario (Canada)

Truong-Huy Nguyen *, Sarah El Outayek, Sun Hee Lim, Van-Thanh-Van Nguyen

[a] Department of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada

ARTICLE INFO

Article history:
Received 2 November 2016
Received in revised form 14 July 2017
Accepted 25 July 2017
Available online 27 July 2017
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Renato Morbidelli, Associate Editor

Keywords:
Extreme rainfall process
Frequency analysis
Statistical modelling
Probability distributions
Extreme-value analysis

ABSTRACT

Many probability distributions have been developed to model the annual maximum rainfall series (AMS). However, there is no general agreement as to which distribution should be used due to the lack of a suitable evaluation method. This paper presents hence a general procedure for assessing systematically the performance of ten commonly used probability distributions in rainfall frequency analyses based on their descriptive as well as predictive abilities. This assessment procedure relies on an extensive set of graphical and numerical performance criteria to identify the most suitable models that could provide the most accurate and most robust extreme rainfall estimates. The proposed systematic assessment approach has been shown to be more efficient and more robust than the traditional model selection method based on only limited goodness-of-fit criteria. To test the feasibility of the proposed procedure, an illustrative application was carried out using 5-min, 1-h, and 24-h annual maximum rainfall data from a network of 21 raingages located in the Ontario region in Canada. Results have indicated that the GEV, GNO, and PE3 models were the best models for describing the distribution of daily and sub-daily annual maximum rainfalls in this region. The GEV distribution, however, was preferred to the GNO and PE3 because it was based on a more solid theoretical basis for representing the distribution of extreme random variables.

1. Introduction

Design and management of various hydraulic structures, particularly urban drainage systems, require information on the probability of annual maximum rainfall occurrence and amount of durations from several minutes to days. This information is often presented in the form of extreme rainfall intensity-duration frequency (IDF) relations (Chow, 1964). In order to construct IDF curves, first, annual maximum rainfall series (AMS) are generally required to perform rainfall frequency analyses due to its much simpler structure comparing to the peak over threshold series (Lang et al., 1999; WMO, 2009a,b). The next step is to select a suitable distribution that could describe well the distribution of the annual maximum rainfall data. This task, however, is not easy and remains as one of the major challenges in engineering practice due to significant spatial and temporal variability of rainfall maxima. In fact, many probability models have been proposed for representing the distribution of annual hydrologic extremes at a single site (Chow, 1964; Kite, 1977; Stedinger et al., 1993; Hosking and Wallis, 1997; Rao and Hamed, 2000; WMO, 2009a; Salinas et al., 2014a,b); however, there is still no general agreement as to which distribution(s) should be used due to the lack of a suitable evaluation procedure. The national guidelines of different countries recommend the use of different distributions. For instance, Log-Pearson 3 has been recommended in the US in Bulletin 17B (Griffis and Stedinger, 2007). The generalized extreme value (GEV) distribution and LP3 are recommended in Australia (Ball et al., 2016). GEV distribution is also a recommended choice in many other countries in Europe, including Austria, Germany, Italy, and Spain (Salinas et al., 2014a). However, many other distributions have also been used popularly, including the Gumbel (GUM) distribution in Finland and Spain, the generalized Pareto (GPA) distribution in Belgium, the generalized logistic (GLO) distribution in the UK (Salinas et al., 2014a). In Canada, the use of a specific distribution is not compulsory, however, LP3, Log-normal three parameters (LN3), GEV, and GUM have been used popularly (Chow and Watt, 1992; Adamowski et al., 1996; Alila, 1999; Hansen, 2015). Environment Canada currently uses GUM to construct at-site IDF curves for all stations in Canada (Environment Canada, 2014). This distribution is also recommended for the development of rainfall IDF relations by the Canadian Standard Association (CSA, 2012). In general, the common method for

* Corresponding author.
E-mail addresses: huy.nguyen5@mail.mcgill.ca (T.-H. Nguyen), sarah.outayek@mail.mcgill.ca (S. El Outayek), sun.lim@mail.mcgill.ca (S.H. Lim), van.tv.nguyen@mcgill.ca (Van-Thanh-Van Nguyen).

http://dx.doi.org/10.1016/j.jhydrol.2017.07.052
0022-1694/

selecting a proper probability model is mainly based on the best fit of the model to the observed data; that is, the model with its best descriptive ability (Mielke and Johnson, 1974; Wilks, 1993; Nguyen et al., 2002a; Laio et al., 2009; Haddad and Rahman, 2011). Consequently, the best-fit selection approach depends strongly on the characteristics of the existing rainfall record at a given site. However, this approach cannot be used to assess the performance of the selected models for extreme events that occur outside the considered rainfall record; that is, based on the best model predictive ability. This characteristic (i.e. model extrapolation or prediction) is considered vital when comparing the performance of different probability models for annual extreme rainfall series, however, there are only a few few studies concerning this point (Wilks, 1993; Öztekin, 2007). Nevertheless, none of these publications have addressed the comparison of a large number of popular probability models for a wide range of short-to-long duration AMS data based on both the descriptive and predictive performance.

In view of the above-mentioned issues, the present study proposes therefore a systematic procedure for assessing and comparing the performance of different probability models in terms of both their descriptive and predictive abilities in order to determine the "best" model that could provide the most accurate extreme rainfall estimates. More specifically, ten common probability distributions for extreme rainfalls were considered in this comparative study (WMO 2009a): Beta-K (BEK), Beta-P (BEP), GEV, GLO, LN3 or Generalized Normal (GNO), GPA, GUM, LP3, Pearson Type III (PE3), and Wakeby (WAK). Graphical and numerical comparison criteria were utilized to evaluate the performance of the selected probability models based on their degree of overall fit to the data, their degree of fit at the right-tails, the accuracy of their right-tail extrapolations, which is of particular importance for engineering design purposes (El Adlouni et al., 2008), and their overall computational facility. The feasibility of the suggested procedure was tested using a total of 63 available AMS data for 5-min, 1-h, and 24-h durations from a network of 21 raingages located in the Ontario region in Canada. These data are provided in Section 2, while the methodology – the systematic approach, is described in details in Section 3. Section 4 presents the results and Section 5 provides the conclusions.

2. Study site and data

A total of 63 annual maximum rainfall series for three different durations from a network of 21 stations located in the Ontario province in Canada were selected for this study as shown in Fig. 1. The record lengths for these datasets vary from 40 years to 75 years. These data were obtained from the website of the Government of Canada (Environment Canada, 2014). Selection of the stations relied on the quality of the data, the adequate length of available historical extreme rainfall records, and the representative spatial distribution of raingages. In order to ensure the quality of data, only data from recording raingages under the supervision of the Atmospheric Environmental Service of Environment Canada were used. At least 40 years of historical records is required in order to provide reliable estimates of rainfall quantiles for the descriptive ability test. Furthermore, half of the sample have at least 20 years of record for the purpose of distribution fitting and then extrapo-

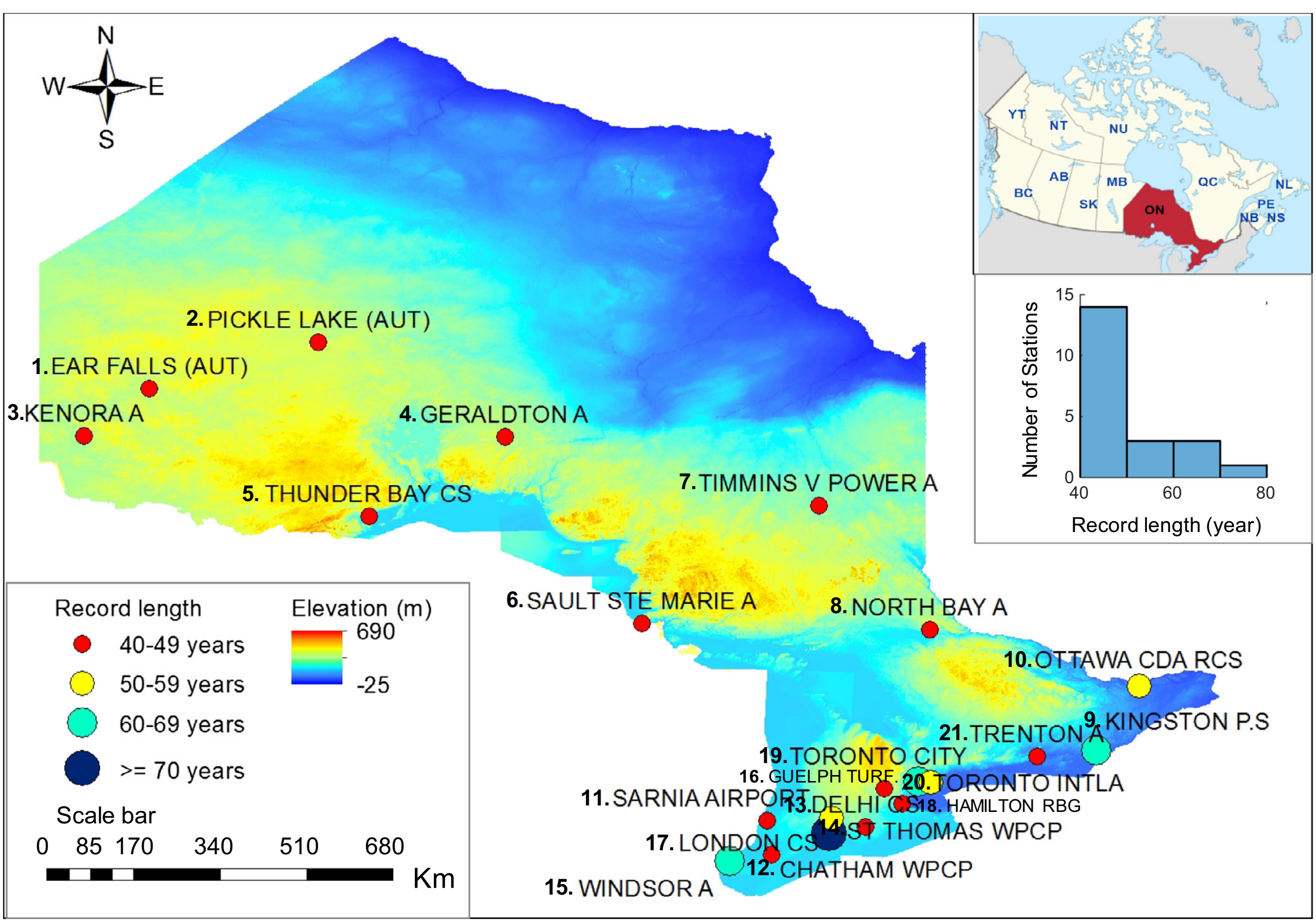

Fig. 1. Locations of 21 study raingages in Ontario. The provincial digital elevation model was obtained from LIO (2016).