Journal of Hydrology 545 (2017) 251–262

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Flood frequency analysis using multi-objective optimization based interval estimation approach

K.S. Kasiviswanathan, Jianxun He *, Joo-Hwa Tay

Department of Civil Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

A R T I C L E I N F O

Article history:
Received 23 December 2015
Received in revised form 29 September 2016
Accepted 12 December 2016
Available online 23 December 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Felix Frances, Associate Editor

Keywords:
Flood frequency analysis
Ensemble simulation
Multi-objective optimization
Prediction interval
Uncertainty

A B S T R A C T

Flood frequency analysis (FFA) is a necessary tool for water resources management and water infrastructure design. Owing to the existence of variability in sample representation, distribution selection, and distribution parameter estimation, flood quantile estimation is subjected to various levels of uncertainty, which is not negligible and avoidable. Hence, alternative methods to the conventional approach of FFA are desired for quantifying the uncertainty such as in the form of prediction interval. The primary focus of the paper was to develop a novel approach to quantify and optimize the prediction interval resulted from the non-stationarity of data set, which is reflected in the distribution parameters estimated, in FFA. This paper proposed the combination of the multi-objective optimization approach and the ensemble simulation technique to determine the optimal perturbations of distribution parameters for constructing the prediction interval of flood quantiles in FFA. To demonstrate the proposed approach, annual maximum daily flow data collected from two gauge stations on the Bow River, Alberta, Canada, were used. The results suggest that the proposed method can successfully capture the uncertainty in quantile estimates qualitatively using the prediction interval, as the number of observations falling within the constructed prediction interval is approximately maximized while the prediction interval is minimized.

1. Introduction

Flooding has always been one of the leading causes of damage among all natural disasters. During past decades, annual losses from floods have reached tens of billions of US dollars and thousands of people were killed each year (Hirabayashi et al., 2013). The major factors causing floods include nature (e.g., extreme rainfall) as well as anthropogenic activities such as deforestation and rapid urbanisation, both of which would increase flood frequency and severity. While the occurrence of the hydrologic extreme events cannot be avoided, proper structural and non-structural measures can effectively reduce the risk of economic losses, environmental damages and social vulnerability. Hydraulic structures, which are commonly designed using a risk-based approach, might fail as the risk level is not always constant over the life span of the structures (Prakash et al., 2014). However, the risk level and its corresponding flood quantile have often been treated as constants in engineering design. Therefore, the hydraulic structures designed based on the *stationary* concept would need to be progressively improved and/or replaced to maintain a required risk level; whereas replacement and improvement of hydraulic structures can be cost-prohibitive in many occasions. On the other hand, more reliable guidelines for design, operation, and management of the structures can be formulated through updating and enhancing non-structural measures (e.g., flood analysis and modeling) as the progress in understanding the physical phenomenon and computational techniques.

Flood frequency analysis (FFA) is one of the non-structural measures and a standard practice in designing water and flood control infrastructure. FFA provides flood quantile estimates, which are used as the basis to design water infrastructure. In practice, flow data (often annual maximum daily flow) are fitted using a theoretical probabilistic distribution function, which is usually selected from a set of several candidate distributions. The commonly used probability distributions in FFA include lognormal, Gamma, Pearson III, generalized extreme value (GEV) and generalized Pareto (GP) distributions. The selection of a particular distribution is largely dependent on the data set(s) used and thus is site- or region-specific. To estimate the distribution parameters, different methods such as method of moment (MOM), maximum likelihood estimator (MLE), probability weighted moment (PWM) and linear moment (L-moment) method have been applied. The accuracy of these methods is assessed based on how closely the theoretical

* Corresponding author.
E-mail address: jianhe@ucalgary.ca (J. He).

http://dx.doi.org/10.1016/j.jhydrol.2016.12.025
0022-1694/

distribution matches with the observations (Strupczewski et al., 2002). Among these methods, L-moment method appears to be more reliable as it is less sensitive to sampling variability and measurement error compared to other methods (Hosking, 1990). Please note that each method, however, has its own merits and demerits in terms of the performance and the procedure to estimate the distribution parameters. In addition to the conventional estimators, a variety of other approaches have been applied in FFA and it is worth mentioning a few here such as neuro fuzzy inference system (Shu and Ouarda, 2008; Basu and Srinivas, 2015), artificial neural network (Shu and Burn, 2004; Aziz et al., 2014), and stochastic simulation (Arnaud et al., 2014).

Similar to many hydrologic models and analyses, FFA has often been performed without considering the inherent uncertainty in the analysis. A few researchers, however, have attempted to quantify uncertainty in FFA. For instance, Wood and Rodríguez-Iturbe (1975) applied Bayesian approach, in which normal and lognormal distributions were considered, to assess uncertainty. However, the assumption of a normal/or lognormal distribution is often not valid for FFA. Tang (1980) developed a procedure to account model structure uncertainty through coupling different probability distributions into FFA. Reis and Stedinger (2005) employed Bayes theorem to enhance quantile estimation and quantify distribution parameter uncertainty. It has been reported that Monte Carlo based simulation approaches produce unique distribution which is not generally biased to other types of distribution through large number of simulations (Reis and Stedinger, 2005; Lee and Kim, 2008; Halbert et al., 2016). Most recently, a few studies showed that the bootstrap based sampling method provides reliable estimate of prediction interval in hydrologic frequency analysis (Tung et al., 2014; Hu et al., 2015). To summarize, most of the methods mentioned above for quantifying uncertainty in FFA are computationally challenging due to the fact that they often require probabilistic and statistical information on the parameters of the selected probability distribution; however such an assumption is hard, if not impossible, to be justified. In the literature, the approach of the ensemble of quantile estimates has often been applied to quantify the uncertainty. Based upon the uncertainty source(s) of interest, the quantification of uncertainty can be carried out using the ensemble simulation approach through varying the data window (e.g., random sampling from the data set) (Chowdhury and Stedinger, 1991; Salas et al., 2013; Obeysekera and Salas, 2014), the distribution parameters (Reis and Stedinger, 2005), the selected distribution type (Tang, 1980) as well as the method for estimating distribution parameters.

There is no doubt that both natural and anthropogenic factors, such as changes/variability in land use, water consumption and global climate, alter hydrological processes and in turn hydrological variables (e.g., flow). To enhance the reliability and resilience of water resources systems, water management agencies need to work towards improved methods for better incorporating the uncertainty introduced by the changes. On the other hand, wide prediction interval, which reflects high level of uncertainty, produced in the modeling/analysis approaches would, however, jeopardize their applications in real world. Therefore approaches, which can provide optimal/tight prediction interval, are always desirable. For FFA, in contrast to the existing methods for quantifying uncertainty, this paper proposed a novel approach to quantify the optimal prediction interval for quantile estimates. This was carried out using the combination of an ensemble simulation (Latin hypercube sampling (LHS) used for sampling) and a multi-objective optimization solved by a search algorithm, called Genetic Algorithm (GA). The advantage of using the proposed method for FFA lies in that it can define the optimal variability of distribution parameters based on the data used in the analysis, so as to ensure the reliability/precision of the proposed method in estimating flood quantiles and quantify their corresponding prediction intervals.

The paper is organized as follows. Followed by the introduction, the description of the study area and the data sets used is given. In the following section the challenges for using the conventional FFA are illustrated, along with detailed introduction of proposed methodology. Subsequently, the analysis results using the proposed methodology are presented and discussed. The conclusions drawn from the paper are then presented at the end.

2. Study area and data description

The Bow River originates from the Canadian Rockies flowing from west towards east through three geographic regions including the mountains, the foothills, and the prairies. The climate of the Bow River Basin is typical of southern Alberta and characterized by long, cold winters and short summers. The temperature in general varies between −12 °C and 18 °C from winter to summer seasons. The annual precipitation in the upper Bow River ranges from 500 to 700 mm in which half of the amount is in the form of snow. In the Calgary region, which is situated in the lower part of the Bow River Basin, there is slightly less annual precipitation (i.e. 412 mm). Approximately 78% of precipitation is rainfall and the remaining is in the form of snow in Calgary. The Bow River is fed by various water sources including surface runoff from late spring to early summer, snowmelt from early spring to early summer, and groundwater recharge throughout a year, which makes the Bow River a complex system. All these challenge the conventional FFA as high flows can be resulted from different mechanisms (snowmelt, rainfall, and their combination) in addition to the potential impacts of climate variability and/or climate change on flows.

There are several flow gauge stations, which are operated by the Water Survey of Canada of Environment Canada, on the Bow River. The City of Calgary is situated approximate in the middle of the river and is the most populated community within the river basin. Most recently in June of 2013, southern Alberta experienced a record-breaking flood in Alberta's history, which caused billions of dollars in losses and demonstrated the detrimental effects of floods. In the City of Calgary, the flood caused an infrastructure loss of $445 million estimated by the flood risk recovery task force. Flow in the Bow River is consistently low from winter until the subsequent early spring. Flow normally starts to increase in May when the melt of mountain snowpack starts feeding the river. The annual maximum flow often occurs around the late-June due to water contribution from both rainfall and mountain snowmelt and it can also be observed in other months including May and from July to September. In the 2013 flood, the rainfall-runoff was considered to be the primary mechanism causing the extremely high peak flow as over 200 mm heavy rain fell in less than two days in many regions in the river basin, particularly west and southwest of Calgary. This paper used the data collected from two flow gauge stations on the Bow River: the Bow River at Banff (named as Banff station) (05BB001) and the Bow River at Calgary (named as Calgary station) (05BH004), which are located in the upper and middle reaches of the river, respectively (Fig. 1). From the river origin to Banff, there are no hydraulic structures regulating flows; whereas several hydraulic dams located between Banff and Calgary and in the tributaries flowing into the Bow River. However, the dams at the upstream Calgary do not have significant impact on the peak flows considering the sizes of the reservoirs. The rationale behind the selection of these two gauge stations lies in that the both gauge stations have long historical data (from 1909 and 2013 at Banff station and from 1912 to 2013 at Calgary station) and are representative to the upper and middle reaches of the river, respectively.