Journal of Hydrology 547 (2017) 557–574

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Appropriate model selection methods for nonstationary generalized extreme value models

Hanbeen Kim [a], Sooyoung Kim [b], Hongjoon Shin [a], Jun-Haeng Heo [a,*]

[a] School of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea
[b] Hydrometeorological Cooperation Center, Republic of Korea

ARTICLE INFO

Article history:
Received 3 March 2016
Received in revised form 11 November 2016
Accepted 6 February 2017
Available online 11 February 2017
This manuscript was handled by Dr. A. Bardossy, Editor-in-Chief, with the assistance of Felix Frances, Associate Editor

Keywords:
Model selection method
Nonstationary GEV model
Akaike's information criterion
Bayesian information criterion
Likelihood ratio test
Proper model selection ratio

ABSTRACT

Several evidences of hydrologic data series being nonstationary in nature have been found to date. This has resulted in the conduct of many studies in the area of nonstationary frequency analysis. Nonstationary probability distribution models involve parameters that vary over time. Therefore, it is not a straightforward process to apply conventional goodness-of-fit tests to the selection of an appropriate nonstationary probability distribution model. Tests that are generally recommended for such a selection include the Akaike's information criterion (AIC), corrected Akaike's information criterion (AICc), Bayesian information criterion (BIC), and likelihood ratio test (LRT). In this study, the Monte Carlo simulation was performed to compare the performances of these four tests, with regard to nonstationary as well as stationary generalized extreme value (GEV) distributions. Proper model selection ratios and sample sizes were taken into account to evaluate the performances of all the four tests. The BIC demonstrated the best performance with regard to stationary GEV models. In case of nonstationary GEV models, the AIC proved to be better than the other three methods, when relatively small sample sizes were considered. With larger sample sizes, the AIC, BIC, and LRT presented the best performances for GEV models which have nonstationary location and/or scale parameters, respectively. Simulation results were then evaluated by applying all four tests to annual maximum rainfall data of selected sites, as observed by the Korea Meteorological Administration.

1. Introduction

Frequency analysis plays an important role in the hydraulic structure design process as well as in the management of water resources. It utilizes appropriate probability distribution models to estimate hydrologic quantiles. Frequency analysis assumes that data is both independent and stationary, i.e., data and its statistical characteristics do not vary over time. Industrialization and urbanization influence change in climatic conditions, and this has caused hydrologic and meteorological data to become nonstationary (Jain and Lall, 2000, 2001; Katz et al., 2002; Milly et al., 2008; Olsen et al., 1999). For example, statistics such as quantiles of hydrologic data, and parameters of probability models may change over time. However, there has been much controversy over the concept of nonstationarity in water resources management and planning. Milly et al. (2008) asserted that nonstationary probabilistic models should be identified and applied because anthropogenic climate change is affecting the extremes of hydrological variables (e.g., precipitation, streamflow and evapotranspiration). In contrast, several scholars have emphasized that the careless application of nonstationarity could lead to the underestimation of variability, uncertainty and risk (Koutsoyiannis, 2011; Lins and Cohn, 2011; Montanari and Koutsoyiannis, 2014; Serinaldi and Kilsby, 2015). Nonetheless, various studies on nonstationarity for hydrological modeling are still being conducted to predict future events under changing environmental conditions. There have been many studies focusing on nonstationary frequency analysis that primarily takes into account covariates, such as time, temperature, and climate indices. Examples of climate indices are Pacific Decadal Oscillation (PDO), Southern Oscillation Index (SOI), Mediterranean Oscillation Index (MOI), North Atlantic Oscillation (NAO), Sea Level Pressure (SLP), and Sea Surface Temperature (SST). These covariates are used to determine parameters of probability distribution models (Brown et al., 2008; Coles, 2001; Griffis and Stedinger, 2007; Katz et al., 2002; Sugahara et al., 2009; Tramblay et al., 2013; Vasiliades et al., 2015; Wang et al., 2004; Wi et al., 2015).

Extreme value theory is a branch of statistics that focuses on the extreme events and the tail behavior of a distribution. The theory

* Corresponding author.
E-mail address: jhheo@yonsei.ac.kr (J.-H. Heo).

http://dx.doi.org/10.1016/j.jhydrol.2017.02.005
0022-1694/

uses the block maxima approach to derive Extreme Value (EV) distributions, including the Fréchet, Weibull, and Gumbel distributions. The GEV distribution unifies the three abovementioned EV distributions. In nonstationary frequency analysis, nonstationary GEV distributions have been proposed and widely used (Cannon, 2010; Coles, 2001; El Adlouni et al., 2007; Kharin and Zwiers, 2005; Leadbetter et al., 1983; Mailhot et al., 2010; Nadarajah, 2005; Vasiliades et al., 2015; Wang et al., 2004; Wi et al., 2015). The nonstationary GEV models proposed by Nadarajah (2005), Vasiliades et al. (2015), and Wi et al. (2015) have been used to conduct nonstationary frequency analysis of the annual maximum rainfall series, using time as a covariate.

In conventional frequency analysis, the χ^2 test, Kolmogorov–Smirnov (KS) test, Cramér von Mises (CVM) test, probability plot correlation coefficient (PPCC) test, Anderson-Darling test, and modified Anderson-Darling test have been used to examine the goodness-of-fit (GOF) for probability models (Heo et al., 2013). In addition to these GOF tests, model prediction error measured by the bootstrap or cross-validation has been used to select an appropriate probability model (Laio et al., 2009; Smyth, 2000). Burnham and Anderson (2002) and Zucchini (2000) introduced and expounded these techniques for model selection. In nonstationary frequency analysis, however, it is not simple to apply goodness-of-fit tests to nonstationary probability distribution models involving parameters that vary with time since these tests should be performed at each time step. Therefore, many studies alternatively recommend the Akaike's information criterion (AIC), corrected Akaike's information criterion (AICc), and Bayesian information criterion (BIC) for the selection of appropriate nonstationary models (Cannon, 2010; Strupczewski et al., 2001a,b; Sugahara et al., 2009; Villarini et al., 2009, 2010). These criteria are straightforward and allow for selecting an appropriate model if the maximized likelihood is calculated. Strupczewski et al. (2001a,b) used the AIC to select the most efficient model out of several nonstationary flood frequency models. Sugahara et al. (2009) applied the AICc (Hurvich and Tsai, 1995) and the rAICc (Burnham and Anderson, 2004) to select the most efficient model out of four nonstationary generalized Pareto distributions. Villarini et al. (2009, 2010) employed the AIC and the BIC to find the degrees of freedom for the Generalized Additive Models of Location, Scale, and Shape parameters (GAMLSS) in a nonstationary framework. Cannon (2010) identified an appropriate nonstationary GEV model using the AICc and the BIC. Vasiliades et al. (2015) identified an appropriate nonstationary GEV model using the AICc and the BIC.

Alternatively, the Likelihood Ratio Test (LRT) has been used in several studies (Clarke, 2002; El Adlouni et al., 2007; García et al., 2007; Katz, 2013; Kharin and Zwiers, 2005; Mailhot et al., 2010; Nadarajah, 2005; Tramblay et al., 2013; Wang et al., 2013), and has been recommended for the selection of an appropriate nonstationary extreme value model (Coles, 2001). Clarke (2002) proposed the Gumbel distribution, involving time as a covariate, and used Generalized Linear Models (GLMs) to fit trend parameters. The LRT was applied to evaluate the goodness-of-fit for the GLMs. Kharin and Zwiers (2005) also evaluated nonstationary GEV models by performing the LRT. Nadarajah (2005), El Adlouni et al. (2007), and Wang et al. (2013) proposed several nonstationary GEV models, and determined the most efficient one by using the LRT. García et al. (2007) conducted the LRT to draw a comparison between stationary and nonstationary GEV models. Mailhot et al. (2010) employed the LRT to compare the nonstationary Ensemble Members (EM) and Annual Maximum (AM) models. Katz (2013) used the AIC, the BIC, and the LRT to select appropriate nonstationary models. Tramblay et al. (2013) selected an appropriate nonstationary Peaks-Over-Threshold (POT) model with the help of the LRT.

The abovementioned studies are only a few ones that compare various model selection criteria to determine an appropriate nonstationary GEV model. Although Stone (1979) described the fundamental characteristics and comparative performance of the AIC and BIC, no specific standards have been set to determine the best criterion for such a model. Therefore, it is likely that an inappropriate model may be selected under nonstationary conditions, and this makes it necessary to determine the most appropriate criterion. Panagoulia et al. (2014) conducted a simple simulation study to evaluate the performances of the AICc and BIC for nonstationary GEV models. However, their results were limited to specific sample sizes and simulation conditions. To get more general results, this study compares the performances of the AIC, the AICc, the BIC, and the LRT, using the Monte Carlo simulation for various sample sizes as well as location, scale, and shape parameters based on stationary and nonstationary GEV distributions. To evaluate the simulation results, the AIC, AICc, BIC, and LRT were applied to the stationary and nonstationary GEV models fitted to the observed annual maximum rainfall data.

2. Model selection criteria

A number of methods can be applied to select appropriate nonstationary models. Of these, the AIC, the AICc, the BIC, and the LRT have been recommended the most. In this study, these tests were applied to various stationary and nonstationary GEV models.

2.1. Nonstationary GEV models

The GEV distribution is widely used for extreme values and includes location, scale, and shape parameters (Lettenmaier and Burges, 1982). The Probability Density Function (PDF) and the Cumulative Density Function (CDF) of the GEV distribution are represented by Eqs. (1) and (2), respectively.

$$f(x) = \frac{1}{\sigma}\left\{1 + \xi\frac{x-\mu}{\sigma}\right\}^{(-1/\xi)-1} \exp\left[-\left\{1 + \xi\frac{x-\mu}{\sigma}\right\}^{-1/\xi}\right] \tag{1}$$

$$F(x) = \exp\left[-\left\{1 + \xi\frac{x-\mu}{\sigma}\right\}^{-1/\xi}\right] \tag{2}$$

where, μ, $\sigma(>0)$, and ξ are the location, scale, and shape parameters, respectively.

In a nonstationary GEV distribution, the GEV parameters can be expressed as various forms of time-dependent function. In this study, the location and scale parameters are expressed as a linear function of time (t), as represented by Eqs. (3) and (4), respectively. It can simply present the increasing or decreasing trend of the location and scale parameters interrelated with the mean and variance of the observed data.

$$\mu(t) = \mu_0 + \mu_1 t \tag{3}$$

$$\sigma(t) = \exp(\sigma_0 + \sigma_1 t) \tag{4}$$

The location parameter varies linearly with time, whereas the scale parameter varies exponentially with time since it is greater than zero (Coles, 2001). For the GEV model, it is difficult to

Table 1
Applied stationary and nonstationary GEV models.

Model	Location parameter	Scale parameter	Shape parameter
GEV(0,0,0)	μ	σ	ξ
GEV(0,1,0)	μ	$\exp(\sigma_0 + \sigma_1 t)$	ξ
GEV(1,0,0)	$\mu_0 + \mu_1 t$	σ	ξ
GEV(1,1,0)	$\mu_0 + \mu_1 t$	$\exp(\sigma_0 + \sigma_1 t)$	ξ