Proc. IAHS, 371, 65–68, 2015
proc-iahs.net/371/65/2015/
doi:10.5194/piahs-371-65-2015

Comparison of nonstationary generalized logistic models based on Monte Carlo simulation

S. Kim, W. Nam, H. Ahn, T. Kim, and J.-H. Heo

School of Civil and Environmental engineering, Yonsei University, Korea

Correspondence to: J.-H. Heo (jhheo@yonsei.ac.kr)

Received: 24 March 2015 – Accepted: 24 March 2015 – Published: 12 June 2015

Abstract. Recently, the evidences of climate change have been observed in hydrologic data such as rainfall and flow data. The time-dependent characteristics of statistics in hydrologic data are widely defined as nonstationarity. Therefore, various nonstationary GEV and generalized Pareto models have been suggested for frequency analysis of nonstationary annual maximum and POT (peak-over-threshold) data, respectively. However, the alternative models are required for nonstatinoary frequency analysis because of analyzing the complex characteristics of nonstationary data based on climate change. This study proposed the nonstationary generalized logistic model including time-dependent parameters. The parameters of proposed model are estimated using the method of maximum likelihood based on the Newton-Raphson method. In addition, the proposed model is compared by Monte Carlo simulation to investigate the characteristics of models and applicability.

1 Introduction

Recently, various evidences of climate change have been observed in hydrologic data (Jain and Lall, 2000, 2001). The effect of climate change on hydrologic data is appeared as a long-term trend or variability in observed hydrologic data. The time-dependent characteristics of statistics in hydrologic data are widely defined as nonstationarity. For analyzing the time-dependent characteristics of hydrologic data, various approaches for the nonstationary frequency analysis have been studied in the recent years. Among many nonstationary frequency analyses, the nonstationary GEV and generalized Pareto models have been mainly suggested for frequency analysis of nonstationary annual maximum and POT (peak-over-threshold) data, respectively. However, the alternative models are required for nonstatinoary frequency analysis because of analyzing the complex characteristics of nonstationary data based on climate change. For this purpose, this study proposed the nonstationary generalized logistic model including time-dependent parameters. The parameters of proposed model are estimated using the method of maximum likelihood based on the Newton-Raphson method. In addition, the proposed model is evaluated by Monte Carlo simulation to investigate the characteristics of models and applicability compared to the stationary generalized logistic model under various simulation conditions.

2 Development of nonstationary generalized logistic model

2.1 Generalized logistic distribution

The generalized logistic distribution is recommended for the flood frequency analysis in UK by Flood Estimation Handbook (Institute of Hydrology, 1999). The cumulative distribution function (CDF) of the generalized logistic distribution are defined as (Hosking and Wallis, 1997)

$$F(x) = \left[1 + \left\{1 - \frac{\beta}{\alpha}(x-\varepsilon)\right\}^{\frac{1}{\beta}}\right]^{-1} \qquad (1)$$

where ε, α, and β are location, scale, and shape parameters, respectively. The range of variable (x) for negative shape parameter $(\beta < 0)$ is given by $\varepsilon + \frac{\alpha}{\beta} \le x < \infty$, and the variable for positive $(\beta > 0)$ shape parameter takes values within $-\infty < x \le \varepsilon + \frac{\alpha}{\beta}$.

Published by Copernicus Publications on behalf of the International Association of Hydrological Sciences.

Table 1. Stationary and nonstationary generalized logistic models.

Model	Location	Scale	Shape	CDF
GLO(0,0,0)	ε	α	β	$F(x)=\left[1+\left\{1-\frac{\beta}{\alpha}(x-\varepsilon)\right\}^{\frac{1}{\beta}}\right]^{-1}$
GLO(1,0,0)	$\varepsilon_0+\varepsilon_1 t$	α	β	$F(x)=\left[1+\left\{1-\frac{\beta}{\alpha}(x-\varepsilon_0-\varepsilon_1 t)\right\}^{\frac{1}{\beta}}\right]^{-1}$

2.2 Nonstationary generalized logistic model

In this study, the time-dependent location parameter is adopted for developing the nonstationary generalized logistic model. The nonstationary location parameter is linearly corresponded to time variable (e.g., $\varepsilon_0+\varepsilon_1 t$) and denotes GLO(1,0,0). In addition, the stationary generalized logistic model denotes GLO(0,0,0). The stationary and nonstationary generalized logistic model is summarized in Table 1.

The parameters of the GLO(0,0,0) and GLO(1,0,0) models are estimated using the method of maximum likelihood. The method of maximum likelihood developed by Fisher (1922) has been one of the best parameter estimation methods. In the method of maximum likelihood, the parameters are determined by maximizing the likelihood function or the log-likelihood function. The log-likelihood functions for the GLO(0,0,0) and GLO(1,0,0) models are respectively defined by

$$\mathrm{LL}_{000}=-N\ln(\alpha)+\left(\frac{1}{\beta}-1\right)\sum_{i=1}^{N}\ln\left\{1-\frac{\beta(x_i-\varepsilon)}{\alpha}\right\} -2\sum_{i=1}^{N}\ln\left[1+\left\{1-\frac{\beta(x_i-\varepsilon)}{\alpha}\right\}^{\frac{1}{\beta}}\right] \quad (2)$$

$$\mathrm{LL}_{100}=-N\ln(\alpha)+\left(\frac{1}{\beta}-1\right)\sum_{t=1}^{N}\ln\left\{1-\frac{\beta(x_t-\varepsilon_0-\varepsilon_1 t)}{\alpha}\right\} -2\sum_{t=1}^{N}\ln\left[1+\left\{1-\frac{\beta(x_t-\varepsilon_0-\varepsilon_1 t)}{\alpha}\right\}^{\frac{1}{\beta}}\right] \quad (3)$$

In this study, the Newton-Raphson method is employed for estimating the parameters.

2.3 Simulation study

In this study, the simulation experiment is performed to investigate the characteristics of stationary and nonstationary generalized logistic models. In the simulation, the performance of each model is evaluated by the root mean square error (RMSE) between the true and calculated quantiles for the generated data by stationary and nonstationary generalized logistic models. The RMSE is defined by

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{c(i)}-x_{T(i)})^2} \quad (4)$$

where n is the number of data sets generated by each model, $x_{c(i)}$ and $x_{T(i)}$ are the calculated and true quantiles for the ith generated data set, respectively. Simulation study is performed under various conditions such as six sample sizes (i.e., $N=30$, 50, 70, 100, 120, and 150), eight shape parameters (i.e., -0.20, -0.15, -0.10, -0.05, $+0.05$, $+0.10$, $+0.15$, and $+0.20$), and a fixed return period ($T=100$ year) for each sample size. The simulation is repeated until 10 000 sets.

For the generated data by the GLO(0,0,0) model, the simulation results are shown in Fig. 1. Generally, the RMSEs of the GLO(0,0,0) and GLO(1,0,0) models decrease as the sample sizes increase as shown in Fig. 1. In addition, the RMSEs of the GLO(0,0,0) model are generally smaller than those of the GLO(1,0,0) model for various conditions.

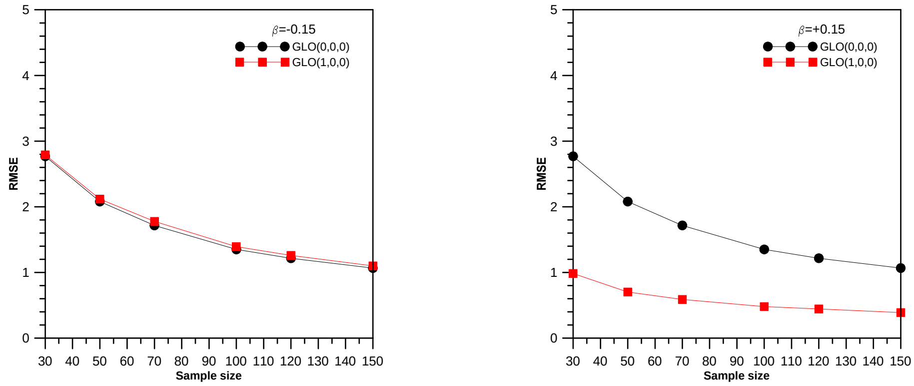
For the generated data by the GLO(1,0,0) model, the simulation results are also shown in Fig. 2.

Generally, all RMSEs of the GLO(1,0,0) model and those of the GLO(0,0,0) model except a case of $\varepsilon_1=-0.01$ and $\beta=+0.20$ decrease as the sample sizes increase. However, the RMSEs of the GLO(0,0,0) model increase as the sample sizes increase for a case of $\varepsilon=-0.01$ and $\beta=+0.20$ as shown in Fig. 2b.

For negative shape parameters, the RMSEs of the GLO(0,0,0) model are smaller than those of the GLO(1,0,0) model until the sample size is 70. Over the sample size of 70, the RMSEs of the GLO(1,0,0) model are smaller than stationary ones. For positive shape parameters, the RMSEs of the GLO(0,0,0) model are smaller than those of the GLO(1,0,0) model until the sample size is 50. Over the sample size of 70, the RMSEs of the GLO(1,0,0) model are smaller than those of the GLO(0,0,0) model. Therefore, the nonstationary generalized logistic model may be applied to relatively long-term data over 70∼100.

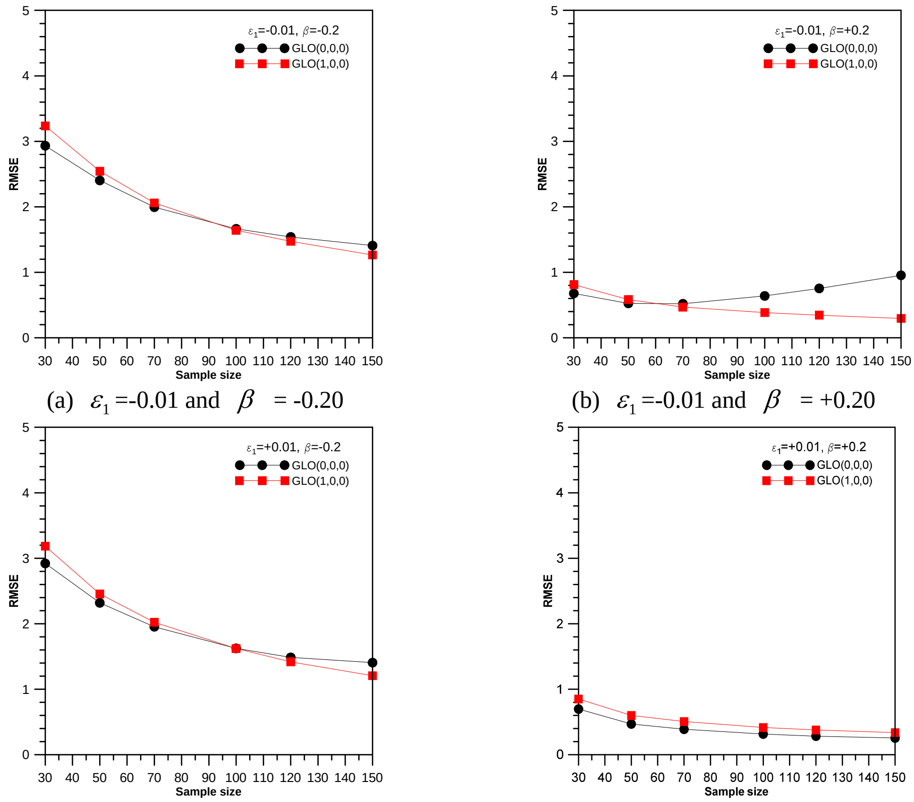
3 Conclusions

For the alternatives of nonstationary frequency analysis, this study proposed a nonstationary generalized logistic model with time-dependent location parameter. The nonstationary parameters are estimated using the method of maximum likelihood. In addition, Monte Carlo simulation is performed to investigate the characteristics of a proposed models and applicability compared to the stationary generalized logistic



(a) β = -0.15 (b) β = +0.15

Figure 1. Simulation results for the generated data by GLO(0,0,0) model.



(a) ε_1 =-0.01 and β = -0.20 (b) ε_1 =-0.01 and β = +0.20
(c) ε_1 =+0.01 and β = -0.20 (d) ε_1 =+0.01 and β = +0.20

Figure 2. Simulation results for the generated data by GLO(1,0,0) model.

model. For the generated data by a stationary model, as the results, the errors of stationary models were generally smallest compared to nonstatnioary model. For the generated data by a nonstationary model, oppositely, the errors of a nonstationary model were smaller than those of a stationary model over 70 or 100 of sample size.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (grant number: 2014006671).

References

Fisher, R. A.: On the mathematical foundations of theoretical statistics, Philos. Trans. Roy. Soc. London Ser. A, 222, 309–368, 1922.

Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: an approach based on L-moments, Cambridge University Press, 1997.

Institute of Hydrology: Flood Estimation Handbook, Willingford, UK, 1999.

Jain, S. and Lall, U.: Magnitude and timing of annual maximum floods: Trends and large-scale climatic associations for the Blacksmith Fork river, Utah, Water Resour. Res., 36, 3641–3651, 2000.

Jain, S. and Lall, U.: Floods in a changing climate: Does the past represent the future?, Water Resour. Res., 37, 3193–3205, 2001.