



The evaluation of regional frequency analyses methods for nonstationary data

W. Nam, S. Kim, H. Kim, K. Joo, and J.-H. Heo

School of Civil and Environmental engineering, Yonsei University, Seoul, Korea

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

Received: 11 March 2015 – Accepted: 11 March 2015 – Published: 12 June 2015

Abstract. Regional frequency analysis is widely used to estimate more reliable quantiles of extreme hydro-meteorological events. The stationarity of data is required for its application. This assumption tends to be violated due to climate change. In this paper, four nonstationary index flood models were used to analyze the nonstationary regional data. Monte Carlo simulation was used to evaluate the performances of these models for the generalized extreme value distribution with linearly time varying location parameter and constant scale and shape parameters. As a results, it was found that the index flood model with time-invariant index flood and time-variant growth curve could yield more statistically efficient quantile when record is long enough to show significant nonstationarity.

1 Introduction

Regional frequency analysis (RFA) has been widely used to estimate more reliable quantiles of extreme hydro-meteorological events. The stationarity of data is required for its application. This assumption tends to be violated due to climate change. Therefore, nonstationary regional frequency analysis (NSRFA), especially nonstationary index flood model (NSIFM), has been recently researched (Cunderlik and Burn, 2003; Cunderlik and Ouarda, 2006; Hanel et al., 2009; Leclerc and Ouarda, 2007; O'Brien and Burn, 2014). The several types of index flood models have been used for NSRFA depending on the time-varying term. There are no prior attempts to evaluate their relative performances.

In this perspective, this paper evaluates the performances of several types of NSIFM through Monte Carlo simulation. Four index flood models including stationary and nonstationary forms are assessed for an arbitrary region with sites which are distributed with nonstationary generalized extreme value (GEV) model.

Section 2 outlines the methodology used in the analysis, including nonstationary index flood models and the GEV distribution. Section 3 discusses the simulation experiment results.

2 Methodology

2.1 Index flood model

Stationary index flood model estimates T -year quantile at site i , Q_T^i by multiplying the index flood (usually sample mean, μ_i) with the regional growth curve (q_T), dimensionless quantile function common to every site in a homogeneous region as given by Eq. (1). This model can be modified to include the nonstationarity of the data. O'Brien and Burn (2014) applied index flood model only with time-varying growth curve (Eq. 1) and Hanel et al. (2009) used the model with time-varying index flood and growth curve (Eq. 2). In addition, the model only with time-varying index flood can also be considered (Eq. 3)

$$Q_T^i = \mu_i \cdot q_T \quad (1)$$

$$Q_T^i(t) = \mu_i \cdot q_T(t) \quad (2)$$

$$Q_T^i(t) = \mu_i(t) \cdot q_T(t) \quad (3)$$

$$Q_T^i(t) = \mu_i(t) \cdot q_T \quad (4)$$

Table 1. RRMSE of 4 NSIFMs for a region with 9 sites (sample size = 50).

Site	Model	Return period (year)								
		2	5	10	20	50	100	200	500	1000
1	1	0.0231	0.0247	0.0274	0.0326	0.0440	0.0561	0.0710	0.0949	0.1159
	2	0.0240	0.0250	0.0287	0.0364	0.0528	0.0697	0.0901	0.1223	0.1506
	3	0.0380	0.0389	0.0426	0.0509	0.0702	0.0906	0.1159	0.1564	0.1924
	4	0.0445	0.0451	0.0468	0.0507	0.0609	0.0731	0.0894	0.1166	0.1414
2	1	0.0233	0.0246	0.0272	0.0322	0.0435	0.0556	0.0705	0.0943	0.1153
	2	0.0228	0.0242	0.0284	0.0366	0.0535	0.0706	0.0911	0.1234	0.1517
	3	0.0368	0.0378	0.0418	0.0504	0.0700	0.0906	0.1160	0.1566	0.1926
	4	0.0464	0.0473	0.0493	0.0535	0.0640	0.0763	0.0926	0.1198	0.1446
3	1	0.0231	0.0249	0.0280	0.0336	0.0454	0.0577	0.0728	0.0968	0.1178
	2	0.0231	0.0250	0.0296	0.0379	0.0549	0.0720	0.0925	0.1248	0.1531
	3	0.0378	0.0388	0.0426	0.0510	0.0702	0.0906	0.1158	0.1562	0.1921
	4	0.0460	0.0471	0.0491	0.0531	0.0633	0.0753	0.0912	0.1180	0.1426
4	1	0.0226	0.0245	0.0278	0.0335	0.0454	0.0578	0.0729	0.0968	0.1179
	2	0.0227	0.0244	0.0288	0.0371	0.0539	0.0709	0.0914	0.1236	0.1518
	3	0.0380	0.0397	0.0441	0.0530	0.0726	0.0931	0.1184	0.1589	0.1948
	4	0.0461	0.0469	0.0484	0.0520	0.0616	0.0732	0.0889	0.1155	0.1399
5	1	0.0233	0.0250	0.0279	0.0334	0.0450	0.0572	0.0722	0.0960	0.1170
	2	0.0236	0.0248	0.0286	0.0362	0.0526	0.0693	0.0895	0.1215	0.1496
	3	0.0364	0.0370	0.0405	0.0487	0.0678	0.0882	0.1134	0.1536	0.1894
	4	0.0445	0.0452	0.0468	0.0505	0.0603	0.0722	0.0880	0.1149	0.1394
6	1	0.0234	0.0248	0.0273	0.0323	0.0434	0.0553	0.0701	0.0938	0.1146
	2	0.0233	0.0251	0.0294	0.0375	0.0541	0.0710	0.0913	0.1233	0.1514
	3	0.0360	0.0372	0.0413	0.0499	0.0692	0.0897	0.1149	0.1552	0.1910
	4	0.0469	0.0476	0.0492	0.0527	0.0622	0.0738	0.0894	0.1160	0.1404
7	1	0.0235	0.0248	0.0275	0.0327	0.0440	0.0560	0.0709	0.0947	0.1155
	2	0.0234	0.0253	0.0296	0.0376	0.0542	0.0710	0.0913	0.1232	0.1513
	3	0.0379	0.0385	0.0419	0.0498	0.0684	0.0885	0.1134	0.1534	0.1890
	4	0.0458	0.0465	0.0479	0.0513	0.0605	0.0720	0.0875	0.1139	0.1381
8	1	0.0249	0.0260	0.0286	0.0336	0.0448	0.0568	0.0716	0.0953	0.1161
	2	0.0232	0.0250	0.0293	0.0374	0.0540	0.0709	0.0911	0.1231	0.1511
	3	0.0377	0.0392	0.0432	0.0517	0.0706	0.0908	0.1157	0.1557	0.1913
	4	0.0457	0.0465	0.0481	0.0517	0.0612	0.0727	0.0882	0.1146	0.1389
9	1	0.0241	0.0254	0.0282	0.0333	0.0446	0.0565	0.0713	0.0949	0.1158
	2	0.0232	0.0252	0.0296	0.0377	0.0542	0.0709	0.0911	0.1230	0.1510
	3	0.0380	0.0397	0.0440	0.0527	0.0719	0.0922	0.1172	0.1573	0.1930
	4	0.0432	0.0442	0.0460	0.0498	0.0597	0.0716	0.0874	0.1141	0.1384

2.2 Nonstationary GEV distribution

Generalized extreme value (GEV) distribution is known to be well fitted to hydrometeorological data. To consider the nonstationarity of data, the parameters can be modeled as a function of time. Thus, the density function of nonstationary GEV distribution is given by

$$F(x) = \exp \left\{ - \left[1 - \beta(t) \frac{x - x_0(t)}{\alpha(t)} \right]^{1/\beta(t)} \right\} \tag{5}$$

where $x_0(t)$, $\alpha(t)$, $\beta(t)$ are location, scale, shape parameters, respectively. In this paper, the scale and shape parameters are assumed as a constant while the location parameter varies lineary with time, which is given by

$$x_0(t) = \varepsilon_0 + \varepsilon_1 \cdot t. \tag{6}$$

Table 2. RRMSE of 4 NSIFMs for a region with 9 sites (sample size = 100).

Site	Model	Return period (year)								
		2	5	10	20	50	100	200	500	1000
1	1	0.0192	0.0194	0.0201	0.0226	0.0297	0.0380	0.0486	0.0657	0.0808
	2	0.0177	0.0186	0.0201	0.0233	0.0304	0.0381	0.0475	0.0624	0.0752
	3	0.0314	0.0321	0.0334	0.0358	0.0416	0.0482	0.0567	0.0707	0.0832
	4	0.0335	0.0342	0.0354	0.0381	0.0449	0.0530	0.0636	0.0813	0.0973
2	1	0.0193	0.0192	0.0198	0.0222	0.0293	0.0377	0.0484	0.0656	0.0807
	2	0.0174	0.0183	0.0201	0.0235	0.0310	0.0388	0.0483	0.0633	0.0762
	3	0.0300	0.0309	0.0325	0.0353	0.0416	0.0485	0.0572	0.0715	0.0842
	4	0.0336	0.0341	0.0352	0.0379	0.0446	0.0527	0.0634	0.0813	0.0974
3	1	0.0200	0.0197	0.0203	0.0227	0.0298	0.0382	0.0489	0.0661	0.0813
	2	0.0172	0.0183	0.0202	0.0236	0.0311	0.0390	0.0486	0.0636	0.0766
	3	0.0306	0.0316	0.0333	0.0362	0.0425	0.0495	0.0583	0.0727	0.0854
	4	0.0330	0.0338	0.0352	0.0380	0.0451	0.0534	0.0642	0.0822	0.0984
4	1	0.0206	0.0200	0.0202	0.0222	0.0290	0.0374	0.0481	0.0654	0.0806
	2	0.0173	0.0180	0.0195	0.0226	0.0298	0.0376	0.0471	0.0621	0.0750
	3	0.0306	0.0315	0.0331	0.0359	0.0421	0.0490	0.0577	0.0721	0.0848
	4	0.0347	0.0352	0.0365	0.0393	0.0463	0.0545	0.0653	0.0833	0.0995
5	1	0.0214	0.0206	0.0207	0.0226	0.0294	0.0377	0.0484	0.0657	0.0809
	2	0.0166	0.0174	0.0193	0.0228	0.0304	0.0383	0.0480	0.0632	0.0762
	3	0.0289	0.0303	0.0323	0.0357	0.0427	0.0501	0.0592	0.0740	0.0870
	4	0.0326	0.0332	0.0347	0.0377	0.0452	0.0539	0.0651	0.0836	0.1001
6	1	0.0223	0.0215	0.0216	0.0235	0.0302	0.0384	0.0491	0.0664	0.0816
	2	0.0166	0.0171	0.0189	0.0224	0.0300	0.0381	0.0478	0.0630	0.0761
	3	0.0293	0.0305	0.0325	0.0358	0.0428	0.0502	0.0594	0.0741	0.0871
	4	0.0334	0.0344	0.0362	0.0396	0.0476	0.0565	0.0678	0.0865	0.1031
7	1	0.0229	0.0218	0.0219	0.0238	0.0305	0.0388	0.0496	0.0670	0.0823
	2	0.0177	0.0183	0.0200	0.0234	0.0309	0.0388	0.0485	0.0636	0.0765
	3	0.0307	0.0319	0.0337	0.0368	0.0435	0.0506	0.0597	0.0743	0.0872
	4	0.0335	0.0344	0.0362	0.0396	0.0476	0.0566	0.0680	0.0868	0.1035
8	1	0.0236	0.0220	0.0216	0.0230	0.0292	0.0373	0.0480	0.0654	0.0807
	2	0.0175	0.0180	0.0197	0.0230	0.0305	0.0385	0.0482	0.0633	0.0763
	3	0.0298	0.0316	0.0340	0.0377	0.0450	0.0526	0.0620	0.0770	0.0901
	4	0.0337	0.0345	0.0361	0.0392	0.0469	0.0556	0.0669	0.0855	0.1022
9	1	0.0246	0.0227	0.0219	0.0230	0.0289	0.0370	0.0477	0.0651	0.0805
	2	0.0183	0.0184	0.0199	0.0232	0.0307	0.0386	0.0483	0.0634	0.0765
	3	0.0315	0.0327	0.0345	0.0375	0.0439	0.0509	0.0597	0.0741	0.0868
	4	0.0346	0.0350	0.0362	0.0390	0.0461	0.0545	0.0656	0.0841	0.1006

3 Performance evaluation of nonstationary index flood models

For the purposes of this study, four models were assessed through Monte Carlo simulation. Two regions with 9 sites were assumed to consider the effect of record length. A region had 9 sites with record length 50 at each site and the other region had 9 sites with 100 year record length at each site. The frequency distribution at each site was assumed as a nonstationary generalized extreme value (GEV) distribution

where location parameter was linearly changed with time. The intercept of location parameter was set up 50. Slope at each site varied linearly from 0.01 at site 1 to 0.018 at site 9. These values of slope were chosen from trend of location parameters of annual maximum rainfall data in South Korea. In this study, the scale and shape parameters were assumed 5 and -0.2 , respectively.

Simulation results are summarized in Tables 1 and 2 for 50 and 100 year record lengths, respectively. In case of 50 year record length, the RRMSE (relative root mean

squared error) of model 1 was lower than those of the other models in most return periods. Among nonstationary models 2–4, the RRMSE of model 2 was lower than models 3 and 4 in return periods under 200 years while the RRMSE of model 4 was lower than models 2 and 3 in return periods over 200 years.

In case of 100 year record length, model 2 gave the lowest RRMSEs while model 4 yielded the highest ones. It indicated that nonstationary model became more efficient than stationary one as record length increased. A possible explanation may be that relatively long record shows more significant nonstationarity.

4 Conclusions

Stationary and nonstationary index flood models were evaluated herein through Monte Carlo simulation for a region with sites which were distributed with nonstationary GEV. The simulation results indicated that index flood model with time-invariant index flood and time-variant growth curve could yield more statistically efficient quantile when record is long enough to show significant nonstationarity.

Future work will focus on the simulation experiments for regions with positive and/or negative trends for the location and scale parameters and more complicated nonstationary GEV distributions.

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

- Cunderlik, J. M. and Burn, D. H.: Non-stationary pooled flood frequency analysis, *J. Hydrol.*, 276, 210–223, 2003.
- Cunderlik, J. M. and Ouarda, T.: Regional flood-duration-frequency modeling in the changing environment, *J. Hydrol.*, 318, 276–291, 2006.
- Hanel, M., Buishand, T. A., and Ferro, C. A. T.: A nonstationary index flood model for precipitation extremes in transient regional climate model simulations, *J. Geophys. Res.*, 114, D15107, doi:10.1029/2009JD011712, 2009.
- Leclerc, M. and Ouarda, T.: Non-stationary regional flood frequency analysis at ungauged sites, *J. Hydrol.*, 343, 254–265, 2007.
- O'Brien, N. L. and Burn, D. H.: A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow, *J. Hydrol.*, 519, 2040–2048, 2014.