

MODELING STREAM FLOW EXTREMES UNDER NON-TIME-STATIONARY CONDITIONS

Ci Yang* and David Hill†

Rutgers, The state University of New Jersey
Department of Civil and Environmental Engineering
623 Bowser Rd., Piscataway, NJ 08854 USA

*e-mail: Jessie_yang06@yahoo.com

†e-mail: ecodavid@rci.rutgers.edu, web page: <http://gaia.rutgers.edu/>

Key words: Generalized extreme value, maximum likelihood estimation, extreme value theory, non-stationary time series, flood frequency analysis

Summary. This paper provides information on extreme value modeling of stream flow extremes from three rivers in the New York City Metropolitan Area using generalized extreme value distribution (GEV). GEV parameters including location, scale, and shape are analyzed to get an appropriate minimum window size for stationary conditions, and then the time series analysis of GEV parameters are performed by using this window size.

1 INTRODUCTION

Extreme value theory (EVT) is a probabilistic theory that can interpret the future probabilities of occurrence of extreme events (e.g., extreme stream flow) using the past records². EVT plays an important role in civil engineering design and water resources management, including flood plain management^{7,8} and design of flood control infrastructure. Traditionally, extreme value theory requires the assumption of temporal stationarity^{5,7,11}. This assumption implies that the historical patterns of recurrence of extreme events are static over time.

The generalized extreme value (GEV) distribution has been commonly applied to the analysis of extreme values based on the assumption of time stationarity^{5,6,10}. Under this assumption, the three GEV parameters, including location, scale, and shape, can be inferred by maximum likelihood (ML) methods⁸, using the large number of data available in the historical record. However, due to human-mediated and natural environmental change, the assumption of a stationary climate is no longer appropriate, and thus, hydrologic conditions, which are driven by climatologic inputs, are also expected to vary in time^{10,11}. Thus, extreme value models must be modified to reflect these temporal changes if they are to continue to be useful for design and management.

Previous studies have applied the GEV distribution to stream flow extremes^{2,4,8} and extreme precipitation^{9,11} in hydrology. The results of these studies suggest that the parameters of the GEV distribution may be a function of other features of the environment, called covariates. Those covariates could be different things depending on the data analyzed, for example, covariates

could be time in precipitation extreme data or drainage area or elevation in flood peak data. Recent studies have found that the location and scale parameters show power law behavior when plotted as a function of drainage area^{4,5}, while some studies showed the shape parameter changes linearly in the log-linear domain⁸. Adlouni et al.² designed and compared four different non-time-stationary GEV models by specifying a linear or a quadratic relationship between location parameter, scale parameter and time with shape parameter being constant using annual maximum precipitation.

The non-time-stationary GEV model can be an efficient tool to take into account the dependencies between extreme valued random variables or the temporal evolution of the climate^{9,11}. In hydrology and civil engineering, for example, such a model can be important significance in flood frequency analysis in designing hydraulic structures through the estimation of reoccurrence of extreme events like peak rainfall¹¹. The extreme event frequency can be modeled by GEV and then can be used to define design criteria for the expected lifetime of the infrastructure¹¹. In such cases, the covariate of time is taken into account.

Annual maximum stream flow time series from 3 stations monitoring three rivers with a record of at least 90 years in the New York City Metropolitan Area are used to examine the distributions of the stream flow extremes from a non-time-stationary perspective. The objectives of this study are to 1) fit observational data with generalized extreme value (GEV) distribution models, 2) explore appropriate window size for identifying non-stationary trends, 3) validate the temporal trends of GEV parameters using both observation and synthetic data, and 4) predict future return values of annual extremes.

2 METHODS

This part designed a procedure in modeling non-time-stationary GEV modeling.

2.1 GEV model

The generalized extreme values model is a combination of three families: Gumbel, Fréchet, and Weibull with a cumulative distribution function (CDF)^{1,2}:

$$F(x | \mu, \alpha, \kappa) = \exp \left\{ - \left(1 - \frac{\kappa}{\alpha} (x - \mu) \right)^{1/\kappa} \right\} \quad \kappa \neq 0$$

$$= \exp \left\{ - \exp \left(1 - \frac{(x - \mu)}{\alpha} \right) \right\} \quad \kappa = 0$$

Where, $\mu \in (-\infty, +\infty)$ is the location parameter related to the magnitude of the time series, $\alpha > 0$ is the scale parameter related to the variability of the time series, and $\kappa \in (-\infty, +\infty)$ is the shape parameter that is related to the heaviness of the tail of the extreme value distribution^{1, 2}. When $\kappa = 0$, the distribution is Gumbel, when $\kappa < 0$, it is Fréchet, and when $\kappa > 0$, it is Weibull. In some studies^{9,11}, $\xi = -\kappa$ is used instead of κ in the GEV CDF equation, however, in hydrology, the parameter κ is more common.

Two dominant methods currently used to estimate the three parameters of the GEV

distribution are maximum likelihood (ML) method, and L-Moments (LMOM). Some studies used only the ML method², some used the LMOM⁵, and some combined two methods^{3,9}. The ML method will result in estimated quantiles of the distribution with small variance³, but may have large variance for small sample sizes⁹. The LMOM method has a relatively large bias compared to the ML method³. One study has shown that either method is qualitatively the same¹. In this study, we will use ML method and we will focus on model parameters that vary as a function of time.

Data: This paper considers three time series (minimum of 90 years) of annual maximum stream flow from United States Geological Survey (USGS) gage stations located at Raritan river (USGS 01400500 Raritan River at Manville NJ), Delaware river (USGS 01463500 Delaware River at Trenton NJ), and Passaic river (USGS 01389500 Passaic River at Little Falls NJ). The stations are selected from three different rivers to explore the generality of our findings beyond a single location, although, all three rivers experience similar climactic conditions. In the following analysis, the river's names are used to represent the gage station located at corresponding river, e.g. Raritan for gage station USGS 01400500, Delaware for USGS 01463500, and Passaic for USGS 01389500. The stationarity of three complete time series is validated by calculating the autocorrelation function (ACF) values. The ACFs for lags >0 are statistically zero for all three time series indicating there is no trend for all three time series.

GEV model for complete dataset: The complete annual stream flow maxima for three gage stations are modeled using GEV distributions assuming stationarity. The Q-Q plots and return level plots (Figure 1) between the observed data and the GEV model show good correspondence. There are, however, a few outliers at the very beginning of lower tail and upper tail. The standard errors of the estimation of the shape parameter for three stations (Raritan, Delaware, Passaic) are 0.07, 0.08, and 0.09 respectively, which is also indicating a goodness of fit of GEV distribution on all stations.

2.2 Window size selection

In order to identify the existence of non-stationary time series and determine the appropriate GEV model parameters, we have to be able to identify the trend in the parameters. Previous studies have used 30^{1,3,10}, 50¹, and 120³ year moving windows to estimate GEV model parameter as a function of time; however, no guidance is available for identifying an optimal window size. For this purpose, we propose a suite of statistical tests that can identify an appropriate window size that is sufficiently large to permit reliable estimation of the GEV parameters, while being sufficiently small to capture the time dependent trend in the model parameters. These tests are applied within nested windows to identify the minimum window size for reliable parameter estimation which gives similar parameters to those estimated for the next larger window in which it is nested. This latter condition ensures smooth variation in the parameter trends between window sizes. In this paper 4 window sizes are used: 10-, 20-, 30-, and 50-year. 10-year samples are constructed by dividing the time series into 10-year non-overlapping periods. Then, 20-, 30-, and 50-year periods are defined such that they contain two, three, and five of the 10-year periods, respectively. The remainder of this section describes the 3 statistical tests in detail.

T-test: The NULL hypothesis of t-test assumes two variables/samples come from the same distribution. The p-value is the significance indicator of a t-test. If the p-value is less than 0.05, we will reject the NULL hypothesis (i.e. the variables are from different distributions). We use the t-test to compare the GEV parameters determined from different window sizes. We expect that if the process is not stationary then the parameters from widely varying window sizes should be statistically different, while those from more similarly sized windows should be statistically indistinguishable. The window size at which the variables become different indicates the largest period over which stationarity can be assumed in the timeseries.

Q-Q test 1: This test determines whether the distribution of GEV models derived from multi-level window sizes have a good correspondence with respect to each other. In this test, quantile-quantile (Q-Q) comparisons of the modeled GEVs from different window sizes are evaluated for linear trend. A linear trend in a Q-Q comparison indicates that the two distributions have the same shape, although not necessarily the same parameters. The metric used to identify linearity between Q-Q pairs is the Pearson's correlation coefficient (r), which measures the strength of the linear trend between the quantiles. This metric indicates how well the GEV model from one window size might be constructed from another distribution using GEV modeled parameters from different window size. The closer the absolute value of this metric is to unity to more similarity between the distributions.

Q-Q test 2: This test compares the quantiles of the modeled data with the observed data to indicate how well the model represents the sample. Again, r is used to indicate goodness of fit between the modeled and observed data within one window size.

3. TREND ANALYSIS

We will use the suite of tests above to evaluate three synthetic and three real-world timeseries' for stationarity. The synthetic time series' are used to validate the test suite on data in which a known temporal trend exists. The real-world time series' are used to evaluate the types of trends that occur in stream flow timeseries.

3.1 Trend on synthetic data

Three sets of synthetic data were generated by sampling 1000 points from a GEV in which one of the parameters (μ , α , and κ , respectively) changed as a function of the time index. These three sets include Model (1): $\mu=8000+1*\text{time}$, $\alpha=3000$, $\kappa=-0.1$; Model (2): $\mu=8000$, $\alpha=3000+1*\text{time}$, $\kappa=-0.1$; Model (3): $\mu=8000$, $\alpha=3000$, $\kappa=-0.5+0.001*\text{time}$. The results of applying Q-Q test 2 on the three synthetic timeseries are shown in Figure 1. As can be seen in this figure, the variance of the correlation coefficient decreases significantly when the window size increases from 10 to 20 years, whereas the distribution of the correlation coefficient for the 20, 30, and 50-year windows appears to be approximately the same. This indicates that 10 samples of the distribution are too few to reliably estimate the GEV parameters, whereas 20 or more samples result in similar goodness-of-fit between the modeled distribution and the observed data. The T-test results (Table 1) shows that the parameters identified using a 20-year

window are statistically indistinguishable from that using a 50-year window, indicating that the stationary period for the synthetic data may be longer than 20 years, and that if the parameters exhibit a temporal trend, it is not significant enough to change the modeled distribution over a 50-year period.

Model (1)														
κ	10-yr	20-yr	30-yr	50-yr	μ	10-yr	20-yr	30-yr	50-yr	α	10-yr	20-yr	30-yr	50-yr
10-yr	1.00	0.42	0.38	0.39	10-yr	1.00	0.30	0.20	0.11	10-yr	1.00	0.74	0.93	0.72
20-yr	0.42	1.00	0.88	0.96	20-yr	0.30	1.00	0.81	0.53	20-yr	0.74	1.00	0.47	0.21
30-yr	0.38	0.88	1.00	0.86	30-yr	0.20	0.81	1.00	0.68	30-yr	0.93	0.47	1.00	0.55
50-yr	0.39	0.96	0.86	1.00	50-yr	0.11	0.53	0.68	1.00	50-yr	0.72	0.21	0.55	1.00

Model (2)														
κ	10-yr	20-yr	30-yr	50-yr	μ	10-yr	20-yr	30-yr	50-yr	α	10-yr	20-yr	30-yr	50-yr
10-yr	1.00	0.64	0.44	0.39	10-yr	1.00	0.74	0.42	0.46	10-yr	1.00	0.64	0.44	0.16
20-yr	0.64	1.00	0.63	0.51	20-yr	0.74	1.00	0.57	0.64	20-yr	0.64	1.00	0.71	0.21
30-yr	0.44	0.63	1.00	0.85	30-yr	0.42	0.57	1.00	0.85	30-yr	0.44	0.71	1.00	0.31
50-yr	0.39	0.51	0.85	1.00	50-yr	0.46	0.64	0.85	1.00	50-yr	0.16	0.21	0.31	1.00

Model (3)														
κ	10-yr	20-yr	30-yr	50-yr	μ	10-yr	20-yr	30-yr	50-yr	α	10-yr	20-yr	30-yr	50-yr
10-yr	1.00	0.19	0.25	0.11	10-yr	1.00	0.80	0.58	0.46	10-yr	1.00	0.82	0.48	0.86
20-yr	0.19	1.00	0.77	0.63	20-yr	0.80	1.00	0.74	0.25	20-yr	0.82	1.00	0.22	0.91
30-yr	0.25	0.77	1.00	0.42	30-yr	0.58	0.74	1.00	0.14	30-yr	0.48	0.22	1.00	0.19
50-yr	0.11	0.63	0.42	1.00	50-yr	0.46	0.25	0.14	1.00	50-yr	0.86	0.91	0.19	1.00

Table 1: P-value (2-side) from Paired T-test for all three parameters (κ : shape, μ : location, and α : scale)

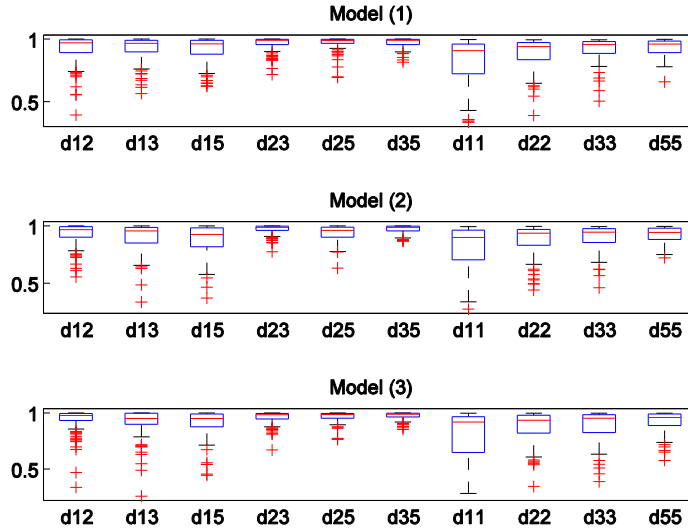


Figure 1: Boxplots of correlation coefficients from goodness of fit test for all three synthetic datasets. ('d12': 10-yr modeled VS 20-yr modeled, 'd13': 10-yr modeled VS 30-yr modeled; 'd15': 10-yr modeled VS 50-yr modeled; 'd23': 20-yr modeled VS 30-yr modeled; 'd25': 20-yr modeled VS 50-yr modeled; 'd35': 30-yr modeled VS 50-yr modeled; 'd11', 'd22', 'd33', 'd55' represent modeled VS observed for each window size, respectively)

The synthetic data sets were partitioned into 99 subsets using the 20-year moving window. GEV parameters were identified for each window and analyzed for temporal trends. As can be seen in Figure 2, the 20-year moving window does capture the temporal trend that was used to generate these data. For model 1, the location parameter increases as a function of time, while the scale and shape parameters remain constant. Similarly, only the scale parameter exhibits a temporal trend for Model 2. However, in Model 3, there is an increasing trend of shape parameter but decreasing trend for location and scale parameters. This indicates a synergistic relationship between the shape and scale parameters that can obscure the true trend from identification by ML methods such as those used in this study. The linear trends identified by our method got very close intercepts and slopes as the true trends designed for the synthetic data sets.

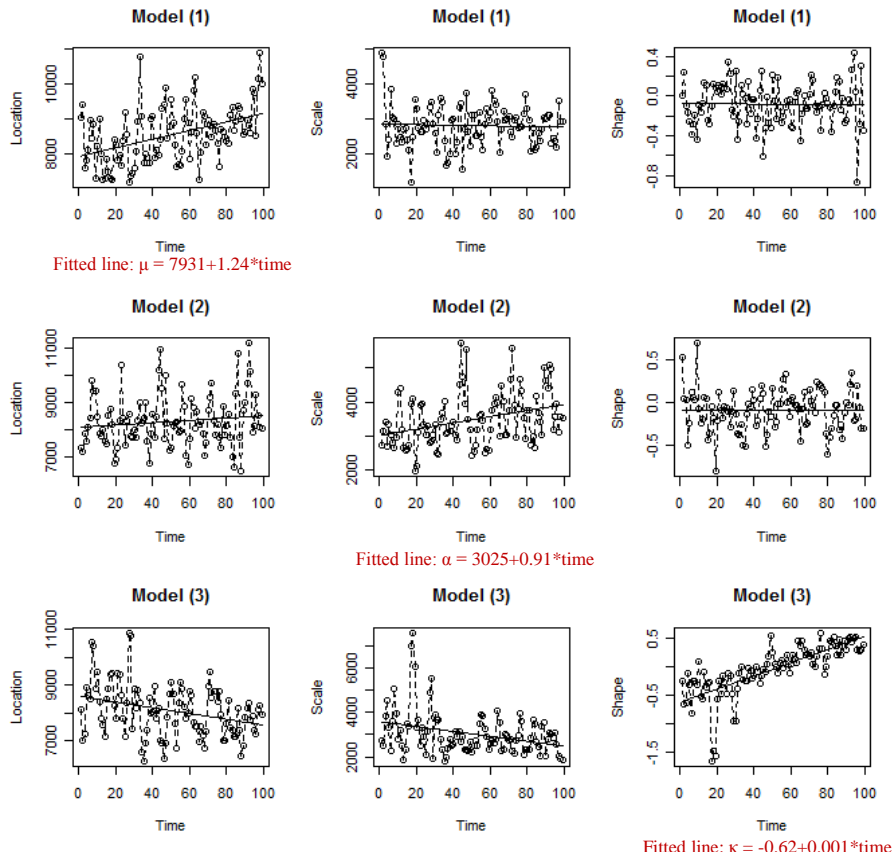


Figure 2: Linear trend test for GEV parameters on synthetic datasets

3.2 Trend on observed data

The window size selected to be 30 following the same methodology described in the method part. The temporal linear trend analysis on parameters is evaluated from observed stream flow in Figure 3. By looking at the p-values for the slope coefficients of the linear regression model, all of them are large than significant level of 0.05, thus we can't say that there is a linear trend

between the parameters and the time. However, in this analysis, we observe similar oscillation patterns for all parameters at all three stations, which may be a random effect because of such small data sample sizes (8~10). To get a larger sample of GEV parameters to prove the non-time-stationarity, we have to get a long period of annual maxima. For instance, a 100 sample of GEV parameters would require more than 1000 years of stream flow time series.

The parameter time series for all three stations are stationary by testing the autocorrelation function (ACF) values to be zero (in the 95% confidence bounds for the NULL hypothesis). Therefore, we are not able to predict any time dependent trend for the parameters.

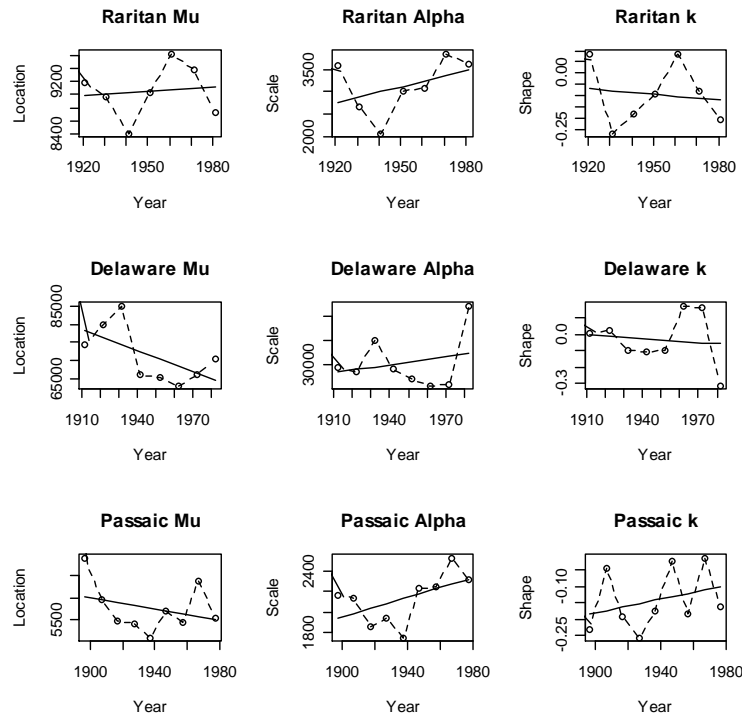


Figure 3: Linear trend test for GEV parameters on observed data (Mu: μ , location parameter; Alpha: α , scale parameter; K: κ , shape parameter)

4 CONCLUSION

GEV model has been successfully applied in our study using the complete data set for all three stations located at Raritan River, Delaware River, and Passaic River. The minimum window size for performing non-time-stationary GEV modeling in this study has been selected to be 30 based on the paired T-test result and goodness of fit using correlation coefficient. Since those three stations are coming from different river with different location, and drainage area, elevation and so on, our optimal window size have applicability on other rivers or other similar types of extreme value such as, the annual peak flood, the annual peak precipitation.

The temporal trend in the analyzing the GEV parameters from 30-year moving window for

our observed data is not statistically significant as the P-value for linear regression models are larger than 0.05, however, we may still observe similar patterns among stations such as oscillations, which may be a result of random effect in the time series.

Because of the lack of long period of field data, we produced non-time-stationary synthetic data by changing the GEV parameters linearly one by one. The optimal window size identified is 20-year. The trend analysis on GEV parameters from synthetic data has shown a very good response in the temporal trend. This validates that window size identified from our methodology works very good. We observed the locations and scale parameters that decrease when the non-time-stationary synthetic data is based on increasing shaper parameter as a linear function of time. We currently don't know why caused that, but it could be explained if we do further analysis by generating more synthetic data. Also, it is suggested to employ a dataset long enough (500 years) to model the temporal trend of the distribution.

At the conference, we plan to discuss how our methods for identifying non-stationarity in the GEV parameters can be used to develop models for the GEV as a function of time that can be used to perform frequency analysis for design under non-stationary conditions.

REFERENCES

- [1] J. E. Morrison, and J. A. Smith, "Stochastic modeling of flood peaks using the generalized extreme value distribution", *Water Resources Res*, **38**, doi:10.1029/2001WR000502, (2002).
- [2] S. El Adlouni et al., "Generalized maximum likelihood estimators for the nonstationary generalized extreme value model", *Water Resour Res*, **43**, W03410, doi:10.1029/2005WR004545, (2007).
- [3] P. Ailliot, C. Thompson, and P. Thomson, "Mixed methods for fitting the GEV distribution", *Water Resour Res*, **47**, W05551, doi:10.1029/2010WR009417, (2011).
- [4] G. Villarini et al., "Analysis of seasonal and annual maximum daily discharge records for central Europe", *J. Hydrol*, **39**, 299-312 (2011).
- [5] G. Villarini and J. A. Smith, "Flood peak distribution for the eastern United States", *Water Resour Res*, **46**, W06504, doi:10.1029/2009WR008395, (2010).
- [6] G. Villarini et al., "Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria", *J. Geophys Res*, **116**, D05103, doi:10.1029/2010JD015038, (2011).
- [7] R. W. Katz, "Statistics of extremes in climate change", *Climate Change*, **100**, 71-76 (2010).
- [8] R. W. Katz et al., "Statistics of extremes in hydrology", *Adv Water Res*, **25**, 1287-1304 (2002).
- [9] V. V. Kharin, and F. W. Zwiers, "Estimating Extremes in Transient Climate Change Simulations", *J. Climate*, **18**, 1156-1173 (2004).
- [10] S. Kao, and A. R. Ganguly, "Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios", *J. Geophys Res*, **116**, D16119, doi:10.1029/2010JD015529, (2011).
- [11] A. Mailhot, and S. Duchesne, "Design criteria of urban drainage infrastructures under climate change", *J. Water Res Plann and Manag*, **136**, 201-208 (2010).