CrossMark

Non-Stationary Frequency Analysis of Extreme Water Level: Application of Annual Maximum Series and Peak-over Threshold Approaches

Ali Razmi[1] · Saeed Golian[1] · Zahra Zahmatkesh[2]

Abstract A great challenge has been appeared on if the assumption of data stationary for flood frequency analysis is justifiable. Results for frequency analysis (FA) could be substantially different if non-stationarity is incorporated in the data analysis. In this study, extreme water levels (annual maximum and daily instantaneous maximum) in a coastal part of New York City were considered for FA. Annual maximum series (AMS) and peak-over threshold (POT) approaches were applied to build data timeseries. The resulted timeseries were checked for potential trend and stationarity using statistical tests including Man-Kendall, Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS). Akaike information criterion (AIC) was utilized to select the most appropriate probability distribution models. Generalized Extreme Value (GEV) distribution and Generalized Pareto Distribution (GPD) were then applied as the probability distribution functions on the selected data based on AMS and POT methods under non-stationary assumption. Two methods of maximum likelihood and penalized maximum likelihood were applied and compared for the estimation of the distributions' parameters. Results showed that by incorporating non-stationarity in FA, design values of extreme water levels were significantly different from those obtained under the assumption of stationarity. Moreover, in the non-stationary FA, consideration of time-dependency for the distribution parameters resulted in a range of variation for design floods. The findings of this study emphasize on the importance of FA under the assumptions of data stationarity and non-stationarity, and taking into account the worst case flooding scenarios for future planning of the

✉ Saeed Golian
s.golian@shahroodut.ac.ir

Ali Razmi
ali.razmi@shahroodut.ac.ir

Zahra Zahmatkesh
Zahra.ZahmatkeshAliabadi@umanitoba.ca

[1] Civil Engineering Department, Shahrood University of Technology, Shahrood, Iran

[2] Faculty of Engineering, Department of Civil Engineering, University of Manitoba, Winnipeg, Canada

watershed against the probable flood events. There is a need to update models developed for stationary flood risk assessment for more robust and resilient hydrologic predictions. Applying non-stationary FA provides an advanced method to extrapolate return levels up to the desired future time perspectives.

Keywords Extreme water levels · Annual maximum series · Peak-over threshold · Frequency analysis · Non-stationarity

1 Introduction

Floods are of the most disastrous natural events, and each year result in substantial environmental, social and economic losses worldwide. Some major examples of flooding over the last decade are floods in the central Europe in 2002 and 2005, and in the East Coast of the United States in 2011 and 2012 (Kron 2005; Zahmatkesh et al. 2015). In general, considering the historical evidences, it can be inferred that the magnitude and frequency of floods have been increased over the last century (Milly et al. 2002; Zahmatkesh et al. 2014). Population and infrastructure rapid growth in flood prone areas, however, have been incentives to enhance understanding of the key drivers that are responsible for those increases. As an example, climate change has been proved to adversely impact the hydrologic cycle (Milly et al. 2002; Goharian et al. 2015; Zahmatkesh et al. 2015).

Considering the vulnerability of coastal areas to extreme water levels (which are representative of coastal floods), regional frequency analysis (FA) is performed to estimate design floods (Karamouz et al. 2014). Frequency analysis is concerned with determining the probability of occurrence of extreme events (Gilroy and McCuen 2012). Utilizing a suitable FA analysis, an optimum design can be achieved (Saf 2008). The exceedance probabilities are generally calculated according to the extreme value analysis of a sample of historical observed data (Mudersbach and Jensen 2010). For hydrologic extremes such as water levels, extreme value theory (EVT) suggests a sound theoretical framework (Hawkes et al. 2008). FA for extreme water levels is usually based on two approaches: annual maximum series (AMS), also called block maxima, and peak-over threshold (POT). While the AMS method fits a probability distribution to the annual maxima of water levels for the defined time intervals (blocks), POT deals with the selection of extreme water levels and modeling all the independent data above a certain threshold (Lang et al. 1999; Khaliq et al. 2006; Silva et al. 2014). Definition of the threshold in this method is critical and depends on the problem of interest (Li et al. 2012).

With the assumptions that the extreme water levels are independent (probability distribution parameters to be estimated come from independent and identically distributed observations), and that the data are stationary (driving forces that affect extreme data do not change in future), the generalized extreme value (GEV) and Generalized Pareto Distribution (GPD) are used as probability distributions for values selected with AMS and POT methods, respectively (Katz et al. 2002; Hawkes et al. 2008; AghaKouchak and Nasrollahi 2010; Li et al. 2015). The POT approach, along with the GPD, has been widely used in the extreme value analyses (Coles 2001; Mackay et al. 2001; Pandey et al. 2001; Ribatet et al. 2007).

Although the assumption of independency for the extreme water levels can be justified by temporal resolution, the likelihood of stationarity, resulted from the existing trends in the mean and/or the variability of the observations, is often suspected (Khaliq et al. 2006). Non-stationarity in the coastal floods generated by severe storms could be attributed to climate

change and variability, land use modification, and watershed regulations, acting separately or together (Katz et al. 2002; Xiong and Guo 2004; Milly et al. 2005, 2008; Gilroy and McCuen 2012; Salas and Obeysekera 2013; Salas and Obeysekera 2014; Vasiliades et al. 2015). Before fitting probability distributions to data, the timeseries requires to be checked against trends for evaluating the presence of non-stationarity (Hawkes et al. 2008; Eregno et al. 2014; Bayazit 2015; Jaiswal et al. 2015). If there is significant non-stationarity in the data, it should be removed to achieve to a stationary timeseries (Salas 1993). However, it is recommended to take into account non-stationarity for FA analysis. One advantage of non-stationary analysis is that the observed data can be used without de-trending. Methods of time-varying moments can be employed for non-stationary flood FA (Strupczewski et al. 2001a, b; Strupczewski and Kaczmarek 2001; Coles 2001; Katz et al. 2002; Khaliq et al. 2006; Villarini et al. 2009; Salas and Obeysekera 2014).

In the non-stationary trend analysis, time-dependent parameters are utilized for the distribution function. Therefore, the results of the extreme value analysis are expected to vary with time. The non-stationary GEV model, for example, is an efficient tool to incorporate the dependencies between the extreme value data (El Adlouni et al. 2007; Cannon 2010). To consider non-stationarity in the GEV model, different expressions have been proposed for the parameters (Tramblay et al. 2013; López and Francés 2013; Ribereau et al. 2008). Generally, the shape parameter is assumed to be constant (Katz et al. 2002; Aissaoui-Fqayeh et al. 2009; Lopez and Frances 2013), while the other two parameters, i.e., location and scale, are considered to be time- or covariant-dependent. Non-stationary distributions have been utilized to overcome the necessity to update assessments by the stationary models and provide more accurate results. Many researchers (e.g., Katz et al. 2002 and Tramblay et al. 2013) have revealed that non-stationary models surpass the classical stationary models in some climate settings. The non-stationary EVT has been significantly used and improved in the last decade (Coles 2001; Khaliq et al. 2006; Mendez et al. 2007; Ribereau et al. 2008).

This paper is aimed to provide a framework for non-stationary analysis of water level for coastal flooding. To do so, historical extreme water levels in a coastal area of New York City (NYC), southern Manhattan, are used. Common statistical methods are applied on the extreme data to detect potential trends. GEV distribution and GPD are described and employed to estimate future design water levels. Various methods for selection of an appropriate threshold to be used in the POT method are also discussed and investigated. Finally, the common design levels are compared with the time-dependent ones.

The structure of the paper is as follows. In the next section, the study area is introduced. Then, the proposed methodology is described followed by providing the results. Finally, a summary and conclusion is given.

2 Case Study

New York has been the largest and richest state in the United States. Manhattan is the most densely populated borough of NYC with the area of 87.46 km^2 and 1.626 million populations in 2013. This borough is surrounded by bays and rives and is threatened by coastal flooding. The most recent recorded events are hurricane Irene and super storm Sandy. Hurricane Irene, occurred in 2011, was a destructive storm with high speed landfalls. Super storm Sandy, happened in 2012, was the largest storm event ever documented in the Atlantic Basin. Based
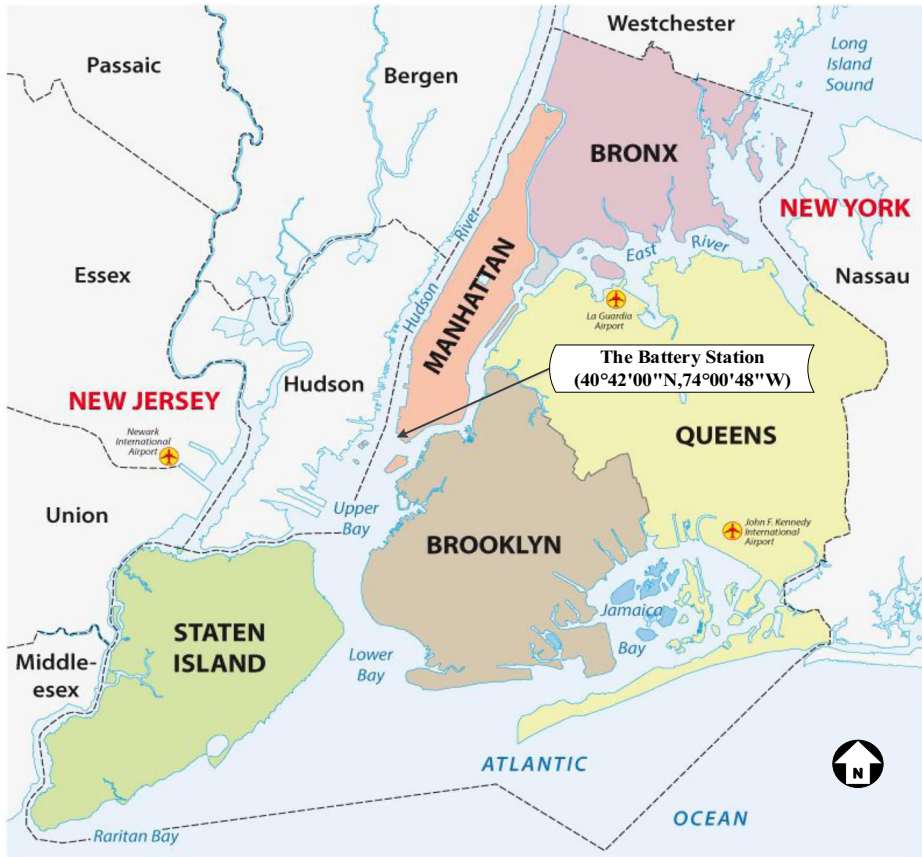
Fig. 1 Location of the Battery Park station in the lower Manhattan (www.shutterstock.com)

on data recorded in the Battery Park station at the southern part of Manhattan (Fig. 1), water level from Sandy at this station was 17.2 ft. (based on the Station Datum (STD)). Trend detection and frequency analysis are performed on water level data from 1920 to 2015 for Manhattan, which are recorded in the Battery Park station.

3 Methodology

FA results are employed to design coastal structures based on a concerned return period or to define flood scenarios for floodplains delineation. For this purpose, analyses are performed on the extreme water level data as representative of coastal flooding. The objective of this study is to examine various probability distribution functions for water level FA considering data non-stationarity. Different distributions are used and water levels for 2, 5, 10, 25, 50, 100, 200, 500, 700, 1000 and 1300-year return periods are estimated. Figure 2 shows the proposed methodology flowchart. The suggested framework includes four main steps: 1- data collection and analysis, 2- extreme value analysis, 3-frequency analysis, and 4-comparing and interpreting the results.
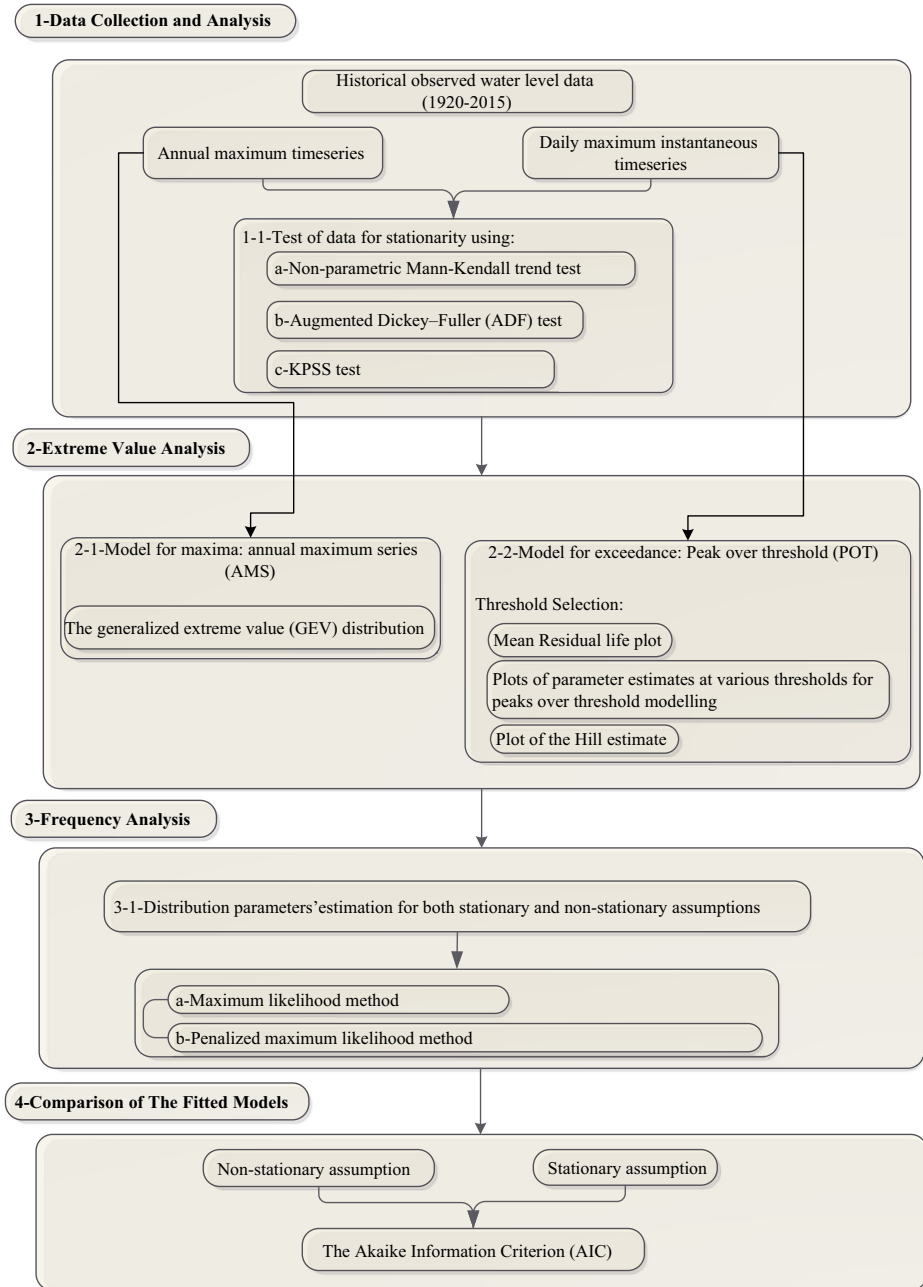
1-Data Collection and Analysis

Historical observed water level data
(1920-2015)

Annual maximum timeseries

Daily maximum instantaneous
timeseries

1-1-Test of data for stationarity using:

a-Non-parametric Mann-Kendall trend test

b-Augmented Dickey–Fuller (ADF) test

c-KPSS test

2-Extreme Value Analysis

2-1-Model for maxima: annual maximum series
(AMS)

The generalized extreme value (GEV) distribution

2-2-Model for exceedance: Peak over threshold (POT)

Threshold Selection:

Mean Residual life plot

Plots of parameter estimates at various thresholds for
peaks over threshold modelling

Plot of the Hill estimate

3-Frequency Analysis

3-1-Distribution parameters'estimation for both stationary and non-stationary assumptions

a-Maximum likelihood method

b-Penalized maximum likelihood method

4-Comparison of The Fitted Models

Non-stationary assumption

Stationary assumption

The Akaike Information Criterion (AIC)

Fig. 2 Flowchart of the proposed methodology for non-stationary frequency analysis of extreme water level

3.1 Step 1 Data Collection and Analysis

Historical instantaneous values of maximum daily and annual water level data for the Battery
Park station were obtained from http://tidesandcurrents.noaa.gov. Using the observed data,

timeseries of: 1- maximum hourly water elevations (one value for each day) named daily maximum instantaneous, and 2- maximum annual water level data (one value for each year), were constructed. Then, the existence of trend in the timeseries of water levels was assessed. The non-parametric Mann-Kendall test (Mann 1945; Kendall 1976) was conducted to examine the variations in the data mean. After confirmation of the trend existence in data, Augmented Dickey–Fuller (ADF) (Said and Dickey 1984) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests were employed to check data stationarity.

3.1.1 Test of Data for Stationary

The Mann–Kendall Test Mann-Kendall (M-K) test is a statistical method utilized for detecting trends in data timeseries. This test is used to identify whether the median of a data timeseries changes over time. In the M-K test, the H0 (null hypothesis) and H1 (alternative hypothesis) correspond to data timeseries without and with trend, respectively. The following relationships are used:

$$s = \sum_{i=1}^{n-1}\sum_{j=1+1}^{n} sgn\left(Y_j - Y_i\right) \tag{1}$$

$$sgn\left(Y_j - Y_i\right) = \begin{cases} +1 & if\ \left(Y_j - Y_i\right) > 0 \\ 0 & if\ \left(Y_j - Y_i\right) = 0 \\ -1 & if\ \left(Y_j - Y_i\right) < 0 \end{cases} \tag{2}$$

$$VAR(s) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} t_p\left(t_p - 1\right)\left(2t_p + 5\right)\right] \tag{3}$$

$$Z_m = \begin{cases} \dfrac{S-1}{\sqrt{VAR(S)}} & if\ S > 0 \\ 0 & if\ S = 0 \\ \dfrac{S+1}{\sqrt{VAR(S)}} & if\ S < 0 \end{cases} \tag{4}$$

where Z_M is the test statistic, n indicates the number of data, Y_i and Y_j are the ith and j[th] observations, q is the number of created clusters with more than two members when the variable (Y) is recorded of several classes, and t_p is the number of data in the pth class. Negative values of Z_M represent a decreasing trend while positive values indicate an increasing trend in the data timeseries. The H0 is rejected when the test statistic is meaningfully different from zero at 5% significance level, i.e., if $|Z_M| > 1.96$, then H0 is rejected which means that trend has been detected in the timeseries.

The Augmented Dickey–Fuller (ADF) Test ADF is utilized to test the stationarity in data. To obtain the coefficients of a model (distribution), this test uses ordinary least squares method.

The significance of the coefficients is estimated through the modified Dickey-Fuller (i.e., t-Student) statistic and compared with the corresponding critical value. When the test statistic is less than the critical value, the H0 is rejected. For variable Y (Y_1, Y_2, ..., Y_n), ADF considers three differential autoregressive (AR) models to detect the presence of a unit root:

$$\Delta Y_t = Y \Delta Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t \tag{5}$$

$$\Delta Y_t = \alpha + \gamma Y \Delta Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t \tag{6}$$

$$\Delta Y_t = \alpha + \beta t + \gamma Y \Delta Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t \tag{7}$$

$$e_t \sim N \left(0, \sigma^2 \right) \tag{8}$$

where Y_t indicates variable at time t, α is an intercept constant, β is the time trend coefficient (the coefficient that is showing the process root), p is the lag order of the AR1 (autoregressive model of order 1) process, and e_t is the residual error. The differences between Eqs. 5–7 indicate the existence of the deterministic drift (α) and linear time trend (βt). The test focus is on determining whether β equals to zero (which means that the timeseries has a unit root) (Said and Dickey 1984).

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test H0 in this test indicates the stationarity of timeseries around mean or a linear trend, while H1 assumes that timeseries is non-stationary because of existence of a unit root. Timeseries Y_1, Y_2, ..., Y_n can be decomposed to a deterministic trend (β_t), random walk (r_t) and stationary error (ε_t):

$$Y_t = r_t + \beta_t + \varepsilon_t \quad t = 1, ..., T \tag{9}$$

$$r_t = r_{t-1} + U_t; \quad U_t \sim N \left(0, \sigma_U^2 \right) \tag{10}$$

For timeseries Y_t with a deterministic stationary trend, H0 is expressed as $\sigma_U^2 = 0$ which means that the intercept is a fixed component, against the alternative of a positive σ_U^2. For this case, the residual error $e_t = \varepsilon_t$ where $e_t = Y_t - \overline{Y}$ and $Y_t = r_0 + \beta_t + \varepsilon_t$ (H0). The H1 states that $e_t = r_t + \varepsilon_t$ which means the process has a unit root. The general form for the KPSS test is as follows:

$$\text{KPPSS} = \frac{1}{T^2} \sum_{t=1}^{t} s_t^2 / \hat{\sigma}_\infty^2 \tag{11}$$

$$s_t = \sum_{j=1}^{t} e_j; \tag{12}$$

$$\hat{\sigma}_\infty^2 = \lim_{t \to \infty} var \left(\sum_{t=1}^{t} r_t \right) \tag{13}$$

The value of the test statistic could be shown by $KPSS_\mu$ (statistic for testing stationarity around mean) or by $KPSS_T$ (statistic for testing stationarity around a trend). If the computed statistic is greater than a critical value at the given level of significance, the H0 (i.e., stationarity) is rejected (Syczewska 2010). Critical values of the KPSS test are provided in Kwiatkowski et al. (1992).

3.2 Step 2 Extreme Value Analysis (EVA)

EVA is a method to study the tail (i.e., extreme) values of a distribution. EVA methods are opposed to the standard approaches, which are usually concerned with the average behavior of data, and intend to capture the behavior of low-frequency high impact events. While the theory of AMS is a solution for modeling extreme values, the POT approach is a technique for modeling violations from a desired high value (the excesses above a big threshold). It should be noted that several distribution functions such as Weibull, Log-normal, GEV and GPD were fitted to data, however, considering previous studies in the study area (e.g., Karamouz et al. 2015), GEV distribution and GPD are used as the probability distributions for AMS and POT methods, respectively.

3.2.1 Model for Maxima: Annual Maximum Series (AMS)

In AMS frequency analysis, annual extreme values are used to obtain the parameters of a probability distribution.

The Generalized Extreme Value (GEV) Distribution The GEV distribution parameters including location (μ), scale (σ), and shape (ξ) specify the center of the distribution, the deviation around mean and the tail behavior of the distribution, respectively. These parameters can be inferred as linear or quadratic approximations of time covariates. In the classic stationary GEV model, all of the parameters are constant. The general form for the non-stationary GEV model is:

$$Y_t \sim GEV(\mu_t, \sigma_t, \xi_t) \tag{14}$$

$$GEV(\mu_t, \sigma_t, \xi_t) = \begin{cases} \exp\left(\left[1 + \xi\left(\frac{Y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0 \\ \exp\left(-e^{-\left(\frac{Y-\mu}{\sigma}\right)}\right) & \text{if } \xi = 0 \end{cases} \tag{15}$$

In the non-stationary GEV, one or more parameters could be functions of time. For example, the location parameter could be a linear function of a co-variable (i.e., $\mu_t = \beta_1 + \beta_2 Y_t$). In another form, the location parameter could be a quadratic function of a temporal co-variable (i.e., $\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2$).

3.2.2 Model for Exceedance: Peak over Threshold (POT)

This method is employed to analyze the distribution of exceedances above a considered high threshold. In POT, the GPD is used as an approximation of the distribution of the tail (extremes). The three-parameter GPD is expressed as:

$$G_{u,\sigma,\xi}(Y) = \begin{cases} 1-(1+\xi/\sigma(Y-u))^{-1/\xi} & if\ \xi \neq 0 \\ 1-e^{-(Y-u)/\sigma} & if\ \xi = 0 \end{cases} \tag{16}$$

where u is the threshold, ξ indicates the shape and σ shows the scale parameter. The shape parameter is associated with the heaviness of the tail. For data observations x_i, $i = 1, \ldots, n$ and threshold u, a GPD is fitted to the exceedances above u through maximum likelihood, and accordingly, the shape and scale parameters are determined. The following equation relates the GEV and GPD models:

$$G_{u,\sigma,\xi}(Y) = 1 + logGEV_{\mu,\sigma,\xi}(Y) \tag{17}$$

One advantage of POT is that it offers control on the number of events related to an interested threshold. For this method, at the first step, a proper threshold should be determined and thereafter, with the given threshold, the GPD can be fitted to data. Parameters of the distribution can be obtained by maximizing the likelihood function (Chen 2014). The following methods are utilized for the estimation of the optimal threshold (Parent and Bernier 2003).

Mean Residual Life Plot If Y has a GPD with scale and location parameters of σ and μ, respectively, then:

$$E(Y) = \frac{\sigma}{1-\mu} \tag{18}$$

Assume that the GPD is a valid model for the excesses of a threshold u_0 generated by timeseries $x_1, x_2, \ldots, x_{n,}$. For random variable x:

$$E(x-u_0|x>u_0) = \frac{\sigma_{u_0}}{1-\mu} \tag{19}$$

where σ_{u_0} indicates the scale parameter corresponding to the excess of the threshold u_0. If the GPD is valid for the excesses of the threshold u_0, it should also be valid for all thresholds $u > u_0$, conditional on the proper change of σ to σ_u. Therefore, for $u > u_0$:

$$E(x-u|x>u) = \frac{\sigma_u}{1-\mu} = \frac{\sigma_{u_0} + \mu u}{1-\mu} \tag{20}$$

For $u > u_0$, $E(x-u|x>u)$ is a linear function of u. Furthermore, $E(x-u|x>u)$ is the mean of the excesses of the threshold u. Based on Eq. 18, these estimates are expected to change linearly with u, at levels of u for which the GPD model is appropriate. This results in:

$$\left\{ \left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}\left(x_{(i)}-u\right) \right) : u < x_{max} \right\} \tag{21}$$

where $x_{(i)}, \ldots, x_{(n_u)}$ includes the n_u observations that exceed u, and x_{max} is the largest amount of the x_i. Eq. 21 establishes the mean residual life plot. Above a threshold u_0 at which the GPD

offers a valid estimate to the excess distribution, the mean residual life plot should be nearly linear in u.

Plots of Parameter Estimates at Various Thresholds for POT Modeling When mean excess function for values greater than a considered threshold is a line with positive slope, it denotes that the data follow GPD with a positive tail. The mean excess function for data with exponential distribution is a horizontal line and for data with low density tail is a line with a negative slope.

Plot of the Hill Estimate Consider there exist some losses (extreme values) from the distribution and the tail coefficient (ξ) needed to be estimated. The following steps should be taken (Makarov 2007):

1- Choosing a threshold $u > 0$.
2- Choosing all losses greater than or equal to u and subtract u from them.
3- Fitting the shifted Pareto distribution to the selected losses by a maximum likelihood estimator.

If $x_1, \ldots, x_n$ denote the losses greater than or equal to u, the probability distribution function for the shifted Pareto distribution is given by:

$$\hat{p}_{\xi,u}(x) = \frac{d}{d_\xi}\hat{p}_{\xi,u}(x) = \frac{d}{d_x}\left(1-\left(1+\frac{x}{u}\right)^{-\frac{1}{\xi}}\right) = 1/\xi u\left(1+\frac{x}{u}\right)^{-\frac{1}{\xi}-1} \tag{22}$$

The log-likelihood function is:

$$\sum_{i=1}^{n} log\left[\hat{p}_{\xi,u}\left(x_i-u\right)\right] = \sum_{i+1}^{n}\left[\left(-\frac{1}{\xi}-1\right)log\left[\frac{x_i}{u}\right]-log[\xi u]\right] \tag{23}$$

Taking the derivative of the log-likelihood function with respect to ξ:

$$\frac{d}{d_\xi}\sum_{i=1}^{n} log\left[\hat{p}_{\xi,u}\left(x_1-u\right)\right] = \sum_{i=1}^{n}\left[\left(\frac{1}{\xi^2}log\left[\frac{x_1}{u}\right]-1/\xi\right)\right] \tag{24}$$

Setting the derivative equal to zero, the estimator for the tail coefficient ξ is obtained as follows:

$$\xi = \frac{1}{n-2}\sum_{i=1}^{n} log[x_i/u] \tag{25}$$

The above estimate is called the Hill's estimator (Hill 1975).

3.3 Step 3 Frequency Analysis

Three packages in R software were used for FA. These packages include extRemes/ismev, POT and texmex. Only the first package can be used for block maxima method; however, the three of them are appropriate for the threshold method. Parameter estimation method for extRemes/ismev is maximum likelihood, and for POT is penalized maximum likelihood. Nevertheless, texmex can use both methods of estimation. These packages have the ability

for non-stationary analysis as well as selecting/changing the values of the distribution model parameters (Gilleland et al. 2013).

ismev package allows for the incorporation of non-stationarity (Gilleland and Katz 2011; Stephenson 2011). The package also comprises of functions for identifying the quality of the fitted distributions as well as functions to assist the selection of an appropriate threshold for the threshold models. extRemes package is an interface for ismev package, but it contains more functionality for estimation of the GEV and GP distributions for the stationary FA. It also provides methods for selection of the threshold. texmex (Southworth and Heffernan 2010) implements the conditional multivariate extreme value modeling method of Heffernan and Tawn (2004). It includes the GPD modeling as well.

3.3.1 Distribution Parameters' Estimation

To incorporate non-stationarity, location and scale parameters were allowed to vary with relevant covariates:

$$\mu(t) = \beta_{o,\mu} + \beta_{1,\mu} Y_1 + \ldots + \beta_{n,\mu} Y_n, \tag{26}$$

$$\sigma(t) = \beta_{o,\sigma} + \beta_{1,\sigma} Y_1 + \ldots + \beta_{n,\sigma} Y_n, \tag{27}$$

where $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients and Y_i 's represent the covariates. The shape parameter is considered to be constant. To estimate the distributions' parameters with both stationary and non-stationary assumptions, likelihood based approaches including maximum likelihood and penalized maximum likelihood are used.

Maximum Likelihood Estimation (MLE) The likelihood is expressed as follows:

$$l(\theta) = \prod_{t=1}^{n} g(Y_t|\theta)_{+} \tag{28}$$

where g is the probability density function of variable Y_t, n is the number of observations and θ indicates the set of the parameters with trend (that could be μ_t, σ_t and ξ_t). The plus sign (+) subscript indicates that the argument should be positive (Obeysekera and Salas 2014). The MLE of θ is the value of θ that maximizes $l(\theta)$.

Penalized Maximum Likelihood Estimation (PMLE) The MLE does not necessarily offer a single solution for the assessment of the parameters of a distribution (Ketterer 2011). MLE usually fails with GPD function with small observed data (Hosking et al. 1985). To acquire a feasible distribution, penalized (modified) likelihood function is used through the application of some functions on the distribution parameters (ξ, is penalized). Quadratic penalization is typically utilized for this purpose. MPLE corrects some of the issues associated with the MLE. In this method, instead of maximizing the log-likelihood $l(\theta)$, the following expression is maximized:

$$l_p(\theta) = l_p(\theta) - Q \tag{29}$$

where Q is a non-negative value. The value of θ that maximizes $l_p(\theta)$ is called a PMLE.

3.4 Step 4 Comparison of the fitted distributions

Akaike Information Criterion (AIC) AIC weights the goodness of fit of a model (Akaike 1974):

$$AIC = -2l + 2K \tag{30}$$

where l is the log-likelihood value estimated for the fitted model and K is the number of the model parameters. Higher ranked models have lower AIC scores.

4 Results

4.1 Test of Stationary for Extreme Water Level Data

Test for stationarity was performed by checking the water level timeseries against a linear trend. The significance of this trend was investigated using the M–K test. The M–K variable τ measures the strength of the relationship between time and water level variables. τ was obtained 0.249, which can be considered significant at 95% significance level. P-value is a function of the observed sample results used for testing a statistical hypothesis. The p-values less than 0.05 indicate a statistically significant improvement in the model performance at the 5% significance level. Since the p-value was smaller than the significance level (α), it is concluded that the null hypothesis of no trend should be rejected and the alternative hypothesis was therefore accepted. The absolute value of Z is 2.5 which is greater than Zα/2 (i.e., 1.96) and p-value (0.0123) is less than the significance level of 0.05.

Timeseries of water level for period 1920–2015 were tested for stationarity using ADF method. The critical values of ADF test in three significance levels were obtained. It was checked if the value of ADF statistics was lower than the critical values, i.e., if the null hypothesis can be accepted. The ADF static test was obtained 0.528, more than all the critical values. Therefore, data timeseries was non-stationary. Furthermore, the p-value was 0.828 which rejected the null hypothesis (i.e., data with no trend).

For KPSS test, in the case of regression with drift (α in Eq. 6), at 1% significance level, i.e., when probability of rejection is less than 1%, data were identified to be non-stationary, while at 5% and 10% significance levels, data were stationary. Since p-value at the significance level of 1% was 0.0192, the non-stationary result was acceptable. In the case of regression with drift and trend, for significance levels of 1% and 5%, data were non-stationary and for significance level of 10% they were stationary. Since the p-value was lower than the significance level of 0.01, the null hypothesis was rejected. To sum up, considering the aforementioned tests, data timeseries were assessed to be non-stationary.

4.2 Peak over Threshold Analysis for Water Level Data

Mean residual life plot for maximum daily water level data is shown in Fig. 3a. The upper and lower dashed lines in this figure establish the confidence intervals at 95% level. We are searching for a point in the figure where the plots start to be linear with upward sloping. It can be observed that the plot starts with a linear behavior, but then indicates considerable
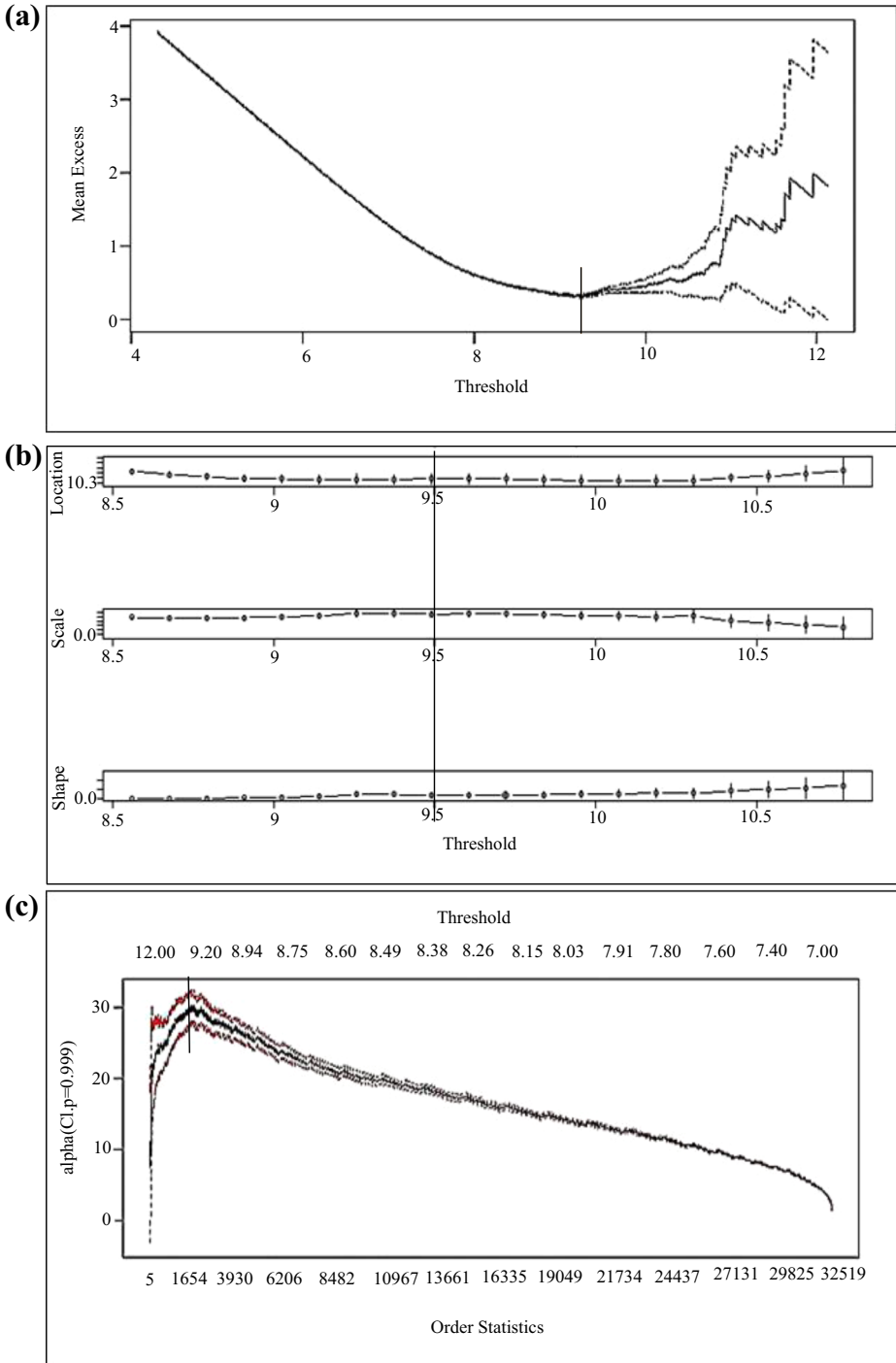
Fig. 3 (a) Mean residual life plot, and (b) location, scale and shape parameter estimates against the threshold with GPD model, and (c) Hill estimate against the number of observations (order statistics) with POT method for maximum daily water level data

curvature in the range 8.5 ft ≤ u ≤ 10 ft. For u > 9.5 ft. the plot is sensibly linear when compared with the confidence intervals. The interpretation of a mean residual life plot is not easy; however, based on the provided explanations, the threshold value of 9.5 ft. was selected to be appropriate.

The parameter life plot can also be utilized to assess the threshold. Figure 3b shows the estimated location, shape and scale parameters for GP distribution against the threshold for maximum water level data. The parameter plot is a plot of maximum likelihood estimates of the parameters of the GPD model. Based on Fig. 3b, selection of u = 9.5 as the threshold is reasonable. Before this value, the plots are nearly horizontal. It should be mentioned that the selection of the threshold is somehow arbitrary. As long as the threshold is within a proper range, in a way that the exceedances above the threshold follow the GPD model, the results can be considered stable.

In Fig. 3c, Hill estimates against the threshold and number of observations for water level timeseries is presented. The Hill estimator in the turning point has relatively a large deviation from the fitted straight line. Moreover, the sequence of the turning points is smaller than $n/10$ in which n equals to the number of observations (Loretan and Phillips 1994). The turning point is the latest sequence of points that fulfill the aforementioned conditions. Based on the Hill estimate, selection of the turning point 9.5 ft. confirms the threshold value identified previously.

4.3 Stationary and Non-stationary Extreme Value Analysis

10 stationary and non-stationary GEV models (GEV0, GEV1, …, GEV9) were fitted to data (Table 1). Table 2 shows the results of frequency analysis based on both annual maxima and peak over threshold methods. Based on this table, GEV9 estimated the maximum values of water levels at different return periods.

It can be seen that in the case of using annual maximum data, for both stationary (GEV0 vs. GEV3) and non-stationary (GEV1 vs. GEV4 and GEV2 vs. GEV5) data, there is not any significant difference between water levels with the same return periods derived from ML and PML estimators. This is the case when using POT data with the assumption of stationarity, i.e. GEV6 vs. GEV8). But, in the case of non-stationarity in POT data (GEV7 vs. GEV9), the

Table 1 Stationary and non-stationary GEV models fitted to the annual water level and daily maximum instantaneous water level data

Model number	GEV index	Model	Number of parameters	Method to estimate parameters
1	0	GEV (μ, σ, ξ)	3	Maximum Likelihood (ML)
2	1	GEV ($\mu = \mu_0 + \mu_1 t$, $\sigma = \exp(\sigma_0 + \sigma_1 t)$, ξ)	5	ML
3	2	GEV ($\mu = \mu_0 + \mu_1 t$, σ, ξ)	4	ML
4	3	GEV (μ, σ, ξ)	3	Penalized Maximum Likelihood (PML)
5	4	GEV ($\mu = \mu_0 + \mu_1 t$, $\sigma = \exp.(\sigma_0 + \sigma_1 t)$, ξ)	5	PML
6	5	GEV ($\mu = \mu_0 + \mu_1 t$, σ, ξ)	4	PML
7	6	GEV (μ, σ, ξ)	3	ML
8	7	GEV ($\mu = \mu_0 + \mu_1 t$, $\sigma = \exp.(\sigma_0 + \sigma_1 t)$, ξ)	5	ML
9	8	GEV (μ, σ, ξ)	3	PML
10	9	GEV ($\mu = \mu_0 + \mu_1 t$, $\sigma = \exp.(\sigma_0 + \sigma_1 t)$, ξ)	5	PML

Table 2 Analysis for water levels in different return periods based on stationary and non-stationary GEV models using AMS and POT methods

Return period	Annual maximum (ft)						Threshold of 9.5 (ft)			
	extRemes model			texmex model			extRemes model		texmex model	
	Stationary GEV (ML estimation)	Non-stationary	Non-stationary	Stationary GEV (PML estimation)	Non-stationary	Non-stationary	stationary	Non-stationary	Stationary	Non-stationary
	GEV0	GEV1	GEV2	GEV3	GEV4	GEV5	GEV6(ML estimation)	GEV7(ML estimation)	GEV8(PML estimation)	GEV9(PML estimation)
2	10.55	11.12	11.03	10.6	11.16	11	9.72	9.80	9.7	9.76
5	11.35	11.93	11.73	11.4	12	11.7	10.06	10.22	10.1	10.1
10	11.93	12.56	12.27	11.9	12.7	12.3	10.43	10.71	10.4	10.5
25	12.74	13.45	13.04	12.7	13.7	13	11.19	11.73	11.2	11.2
50	13.40	14.21	13.70	13.4	14.6	13.7	12.09	12.97	12.1	12.1
100	14.10	15.05	14.42	14.1	15.6	14.4	13.42	14.86	13.5	13.5
200	14.86	15.98	15.32	14.9	16.7	15.2	15.38	17.71	15.6	15.5
300	15.33	16.57	15.74	15.3	17.4	15.7	16.93	20.02	17.2	17.1
500	15.96	17.37	16.43	16	18.4	16.4	19.48	23.87	19.9	19.7
600	16.19	17.67	16.69	16.2	18.8	16.7	20.58	25.56	21.1	20.9
700	16.39	17.93	16.92	16.4	19.1	16.9	21.26	27.14	22.2	22
1000	16.86	18.56	17.46	16.9	19.9	17.5	24.35	31.40	25.2	24.9
1300	17.22	19.05	17.88	17.2	20.6	17.9	26.76	35.18	27.8	27.4

difference between water levels becomes more significant for return periods of greater than 100-years.

4.4 Comparison of the Models

Table 3 shows values of AIC for the proposed distributions. Based on the results, extRemes and texmex GEV models with threshold value of 9.5 ft. (i.e., GEV7 and GEV9 models), with non-stationary location and shape parameters, have the lowest value of AIC and therefore, are the most appropriate distributions for frequency analysis of water level data in the study region. Estimated parameters for these distributions models are shown in Table 3.

Figure 4 compares the results of FA for the selected distribution models. This figure indicates that non-stationary FA with the POT method results in significantly higher values of water level at different return periods. It also shows even with the same assumption (i.e., non-stationarity of parameters) and the same method (POT), application of different methods for parameter estimation (ML vs. PML) could result in different values for water levels in various return periods.

5 Summary and Conclusion

Driving forces such as climate change have been proved to significantly affect the stationarity in the hydrologic data. Non-stationarity cannot be ignored especially in the studies for coastal areas, since it can change the estimated design floods obtained based on the frequency analysis of the observed extreme water level data. In the present study, different stationary and non-stationary distribution models have been fitted to the historical extreme water level data for the lower Manhattan, NYC. In developing the models, two approaches of annual maxima series and peak over threshold were used. AIC criterion was used to compare the models and finally, water level with different return periods was estimated. Using the ADF and KPSS methods as the stationarity tests, it was concluded that data timeseries was non-stationary. Data for FA were selected based on AMS and POT approaches. Then 10 stationary and non-stationary GEV models, were fitted to the selected data. It was shown that in the case of using annual maximum data, for both stationary and non-stationary assumption, except in the case of non-stationarity in POT data, there is not any significant difference between water levels of the same return periods derived from ML and PML estimators. It was also shown that based on

Table 3 AIC test results for distribution models of extreme water levels and estimated parameters for the models with the lowest values of AIC

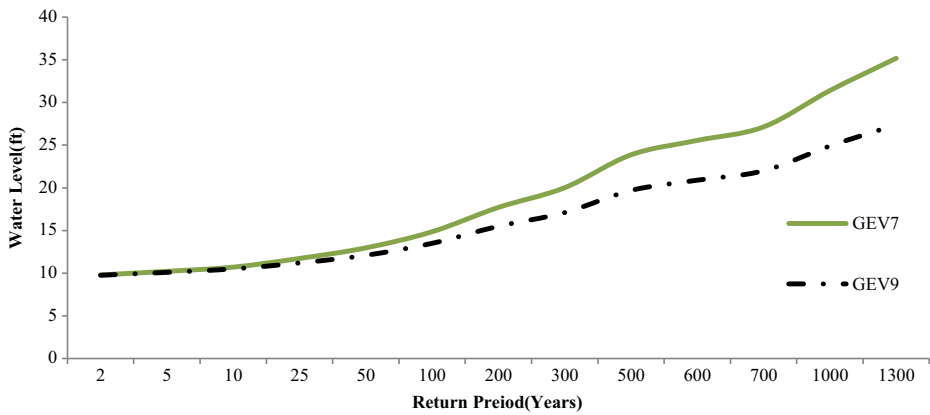| Distribution	Stationary	Non-stationary (location)	Non-stationary (location and scale)
AM (ft) extRemes model	229.698	206.06	206.83
AM(ft) texmex model	229.698	206.067	206.80
Threshold 9.5 (ft) extRemes model	196.09	-	162.14 (μ_0=9.59, μ_1=0.00015, σ_0=0.122, σ_1=0.0001, ξ = 0.60)
Threshold 9.5 (ft) texmex model	196.09	-	171.28 (μ_0=9.78, μ_1=-5.5e-05, σ_0=-1.34, σ_1=2.8e-04, ξ = −2.8e-02)

Fig. 4 Values of extreme water level data in different return periods based on the selected distribution models

AIC criterion, the extRemes and texmex GEV models applied to POT data with threshold value of 9.5 ft. with non-stationary location and shape parameters revealed better performance (the lowest value of AIC) compared to other model and, therefore, are the most appropriate distributions for frequency analysis of water level data in the study region. In general based on the findings, application of different distributions for frequency analysis of extreme water levels, results in obtaining a range of potential values for water level in different return periods. Selecting the type of distribution and the method to choose which data to be analyzed, significantly affects the results. In this study, it was shown that frequency analysis of water level data with the assumptions of stationary and non-stationary results in significantly different results. Moreover, a comparison was made between the distribution parameters' estimation methods to investigate how application of different estimation techniques can affect the FA results. Taking into account non-stationarity due to the natural and/or anthropogenic activities at local or global scales, non-stationary frequency analysis is recommended to be used to estimate values of hydrologic variables in different design periods.

References

AghaKouchak A, Nasrollahi N (2010) Semi-parametric and parametric inference of extreme value models for rainfall data. Water Resour Manag 24:1229–1249

Aissaoui-Fqayeh I, El-Adlouni S, Ouarda TBMJ, St-Hilaire A (2009) Développement du modèle log-normal non-stationnaireetcomparaison avec le modèle GEV non-stationnaire. Hydrol Sci J 54:1141–1156 (In French)

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723

Bayazit M (2015) Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. Environmental Processes 2(3):527–542

Cannon AJ (2010) A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. Hydrol Process 24:673–685. doi:10.1002/hyp.7506

Chen X (2014) Extreme value distribution and peak factor of crosswind response of flexible structures with nonlinear aeroelastic effect. J Struct Eng :04014091. doi:10.1061/(ASCE)ST.1943-541X.0001017

Coles S (2001) An introduction to statistical modelling of extreme values. Springer Series in Statistics. Springer Verlag, London, p 208

El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobee B (2007) Generalized maximum likelihood estimators for the non-stationary generalized extreme value model. Water Resour Res 43:W03410. doi:10.1029/2005WR004545

Eregno FE, Nilsen V, Seidu R, Heistad A (2014) Evaluating the trend and extreme values of faecal indicator organisms in a raw water source: a potential approach for watershed management and optimizing water treatment practice. Environmental Processes 1(3):287–309

Gilleland E, Katz RW (2011) New software to analyze how extremes change over time. Eos 92(2):13–14

Gilleland E, Ribatet M, Stephenson AG (2013) A software review for extreme value analysis. Extremes 16(1): 103–119

Gilroy KL, McCuen RH (2012) A non-stationary flood frequency analysis method to adjust for future climate change and urbanization. J Hydrol 414:40–48

Goharian E, Burian S, Bardsley T, Strong C (2015) Incorporating potential severity into vulnerability assessment of water supply systems under climate change conditions. J Water Resour Plann Manage:04015051. doi:10.1061/(ASCE)WR.1943–5452.0000579

Hawkes PJ, Gonzalez-Marco D, Sánchez-Arcilla A, Prinos P (2008) Best practice for the estimation of extremes: Areview. J Hydraul Res 46:324–332

Heffernan JE, Tawn JA (2004) A conditional approach for multivariate extreme values (with discussion). J R Stat Soc Ser B 66:497–546

Hill BM (1975) A simple general approach to inference about the tail of a distribution. Ann Statist 3:1163–1174

Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27:251–261

Jaiswal RK, Lohani AK, Tiwari HL (2015) Statistical analysis for change detection and trend assessment in climatological parameters. Environmental Processes 2(4):729–749

Karamouz M, Zahmatkesh Z, Nazif S, Razmi A (2014) An evaluation of climate change impacts on Extreme Sea level variability: coastal area of New York City. Water Resour Manag. doi:10.1007/s11269-014-0698-8

Karamouz M, Ahmadvand F, Fereshtehpour M (2015) Flood scenarios determination using nonstationary flood frequency analysis in coastal areas, 9th world congress, Water Resources Management in a Changing World

Katz RW, Parlang MB, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Resour 25:1287–1304

Kendall MG (1976) Rank correlation methods, 4th edn. Griffin

Ketterer F (2011) Penalized likelihood based tests for regime switching inautoregressive models. Ph.D. Thesis, MarburgUniversity, Germany

Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydrometeorological observations: a review. J Hydrol 329:534–552

Kron W (2005) Flood risk = hazard exposure vulnerability. International Water Resources Association, Water International 30:58–68

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit Root. J Econ 54:159–178 North-Holland

Lang M, Ouarda TBMJ, Bobeè B (1999) Towards operational guidelines for over-threshold modeling. J Hydrol 47:103–117

Li F, Bicknell C, Lowry R, Li Y (2012) A comparison of extreme wave analysis methods with 1994–2010 offshore Perth dataset. Coastal Engineering 69:1–11

Li LC, Zhang LP, Xia J, Gippel CJ, Wang RC, Zeng SD (2015) Implications of modelled climate and land cover changes on runoff in the middle route of the south to north water transfer project in China. Water Resour Manag. doi:10.1007/s11269-015-0957-3

López J, Francés F (2013) Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrol Earth Syst Sci 17:3189–3203. doi:10.5194/hess-17-3189-2013

Loretan M, Phillips P (1994) Testing the covariance stationarity of Heavytailed timeseries. J Empir Financ 1:211–248

Mackay EBL, Challenor PG, Bahaj AS (2001) A comparison of estimators for the generalised Pareto distribution. Ocean Eng 38:1338–1346

Makarov M (2007) Applications of exact extreme value theorem. J Oper Risk :115–120

Mann HB (1945) Nonparametric tests against trend. J Econometric Soc Econometrica :245–259

Mendez FJ, Menendez M, Luceno A, Losada IJ (2007) Analyzing monthly extreme sea levels with a time dependent GEV model. J Atm Ocean Technol 24:894–911

Milly P, Wetherald R, Dunne K, Delworth T (2002) Increasing risk of great floods in a changing climate. Nature 415:514–517

Milly PCD, Dunne KA, Vecchia AV (2005) Global pattern of trends in streamflow and water availability in a changing climate. Nature 438(7066):347–350

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: whiter water management? Science 319(5863):573–574

Mudersbach C, Jensen J (2010) Non-stationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline. Journalof Flood RiskManagement 3:52–62

Obeysekera J, Salas JD (2014) Quantifying the uncertainty of design floods under nonstationary conditions. J Hydrol Eng 19(7):1438–1446

Pandey MD, Van Gelder PHAJM, Vrijling JK (2001) The estimation of extreme quantiles of wind velocity using Lmoments in the peaks-over-threshold approach. Struc Saf 23:179–192

Parent E, Bernier J (2003) Bayesian pot modeling for historical data. J Hydrol 274(1–4):95–108. doi:10.1016/S0022-1394(02)00396-7ISSN 0022-1694

Ribatet M, Sauquet E, Grésillon J-M, Ouarda TBMJ (2007) A regional Bayesian POT model for flood frequency analysis. Stoch Environ Res Risk Assess 21:327–339

Ribereau P, Guillou A, Naveau P (2008) Estimating return levels from maxima of non-stationary random sequences using the generalized PWM method. Nonlin Processes Geophys 15:1033–1039

Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71:599–607

Saf B (2008) Application of index procedures to flood frequency analysis in Turkey. Journal of the America Water Resources Association (JAWRA) 44(1):37–47. doi:10.1111/j.1752-1688.2007.00136.x

Salas JD (1993) Analysis and modelling of hydrologic timeseries. In: Maidment DR (ed) Handbook of hydrology. McGraw-Hill Inc., New York, pp 19.1–19.72

Salas J, Obeysekera J (2013) Revisiting the concepts of return period and risk for non-stationary hydrologic extreme events. J Hydrol Eng. doi:10.1061/(ASCE)HE.1943-5584. 0000820

Salas JD, Obeysekera J (2014) Revisiting the concepts of return period and risk for non-stationary hydrologic extreme events. J Hydrol Eng 19:554–568

Silva AT, Portela MM, Naghettini M (2014) On peaks-over-threshold modeling of floods with zero-inflated poisson arrivals under stationarity and nonstationarity. Stoch Environ Res Risk Assess 28(6):1587–1599

Southworth H, Heffernan JE (2010) texmex: Threshold exceedences and multivariate extremes, R package version 1.0

Stephenson AG (2011) ismev: An Introduction to Statistical Modeling of Extreme Values, Original S functions written by Janet E. Heffernan with R port and documentation provided by A. G. Stephenson. R package version 1.35 ed

Strupczewski WG, Kaczmarek Z (2001) Non-stationary approach to at-site flood frequency modeling II. Weighted least squares estimation. J Hydrol 248(1–4):143–151

Strupczewski WG, Singh VP, Feluch W (2001a) Non-stationary approach to at-site flood frequency modeling I. maximum likelihood estimation. J Hydrol 248(1–4):123–142

Strupczewski WG, Singh VP, Mitosek HT (2001b) Non-stationary approach to at-site flood frequency modeling III. Flood analysis of polish rivers. J Hydrol 248(1–4):152–167

Syczewska, E. M. (2010). Empirical power of the Kwiatkowski-Phillips-Schmidt-shin test (No. 45).

Tramblay Y, Neppel L, Carreau J, Kenza N (2013) Non-stationary frequency analysis of heavy rainfall events in southern France. Hydrol Sci J 58:1–15

Vasiliades L, Galiatsatou P, Loukas A (2015) Non-stationary frequency analysis of annual maximum rainfall using climate covariates. Water Resour Manag 29(2):339–358

Villarini G, Serinaldi F, Smith JA, Krajewski WF (2009) On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resour Res 45(8):W08417. doi:10.1029/2008WR007645

Xiong LH, Guo SL (2004) Trend test and change-point detection for the annual discharge series of the Yangtze River at the Yichang hydrological station. Hydrol Sci J 49(1):99–112

Zahmatkesh Z, Karamouz M, Goharian E, Burian S (2014) Analysis of the effects of climate change on urban storm water runoff using statistically downscaled precipitation data and a change factor approach. J Hydrol Eng :05014022. doi:10.1061/(ASCE)HE.1943-5584.0001064

Zahmatkesh Z, Karamouz M, Nazif S (2015) Uncertainty based modeling of rainfall-runoff: combined differential evolution adaptive metropolis (DREAM) and K-means clustering. Adv Water Resour 83:405–420