

## Original Research Article

# Non-Stationary Modeling of Annual Flood Peak Heights of Mahanadi River Basin with The q-Generalized Extreme Value Distribution

## Abstract

In recent years, due to climate change, catastrophic events are increased largely in India. Hence researchers are forced to consider non-stationary flood frequency analysis as an improved method. In this paper, non-stationarity of annual daily maximum flood heights were studied at 12 sites of Mahanadi River Basin (MRB) by analyzing the flood frequency of a stationary model and 4 non-stationary models using time dependent q-GEV model by considering trend as a linear function of its location and scale parameters. The q-GEV distribution is utilized in this study because of its flexibility and accuracy than GEV distribution in modeling extreme flood heights. The results found that there is strong evidence of a linear trend existence for both the location and scale parameters at the Kesinga site; for the location parameter at Pathardi and Simga sites; for the scale parameter at Dharmajagarh, Kotni and Seorinarayan, and no linear trend exists for both location and scale parameters at Alipingal, Bomnidhi, Manendragarh, Mohana, Rajim and Sundargarh, there may be exists other form of trend at these sites. The findings also indicate that nonstationarity is present in the MRB due to climate change, which help to water practitioner for taking precautions against adverse effect of extreme floods.

**Keywords:** Mahanadi River Basin; Non-stationarity; q-GEV Model; MLE.

## 1. Introduction

In the past few years, occurrence of extreme hydrological events such as floods, droughts are increasingly repeated due to impact of climate change, urbanization, deforestation and encroachment of river basin, particularly in India (Singh and Kumar, 2013; Black et al., 2018). In India, each year on an average 1600 persons die as a result of the floods, about Rs. 5600 Crores (73 Million USD) as a damage cost due to floods and 12% of geographic area of India is flood prone (Black et al., 2018). The assumption of stationary is that in a system, statistical characteristics such as mean and variance do not vary over time and show no trend (Sivakumar, 2016). The impact of climate change, global warming, deforestation and urbanization can invalidate the assumption of stationarity (Ray et al., 2019), Gumbel (1941) expressed invalidation of stationary model flood frequency analysis in the situations of climate change and other variations. It is therefore necessary to develop alternative methods that take nonstationarity into account for the effective design of hydraulic control structures. The floods are exacerbated by natural events such as reduced carrying capacity of the water course caused by silting of the river bed, landslide causing obstacles and change in river path and by manmade events such as unplanned urbanization and floodplain encroachment (Singh and Chinnasamy, 2021). The flood frequency analysis of one stationary and 12 non-stationary models were used to evaluate the nonstationarity of Periyar River flow, and their results indicate that climate change and anthropogenic activities are equally responsible for the nonstationarity of the Periyar River discharge (Singh and Chinnasamy, 2021). Seasonality and Trend in natural occurrences such as floods are common causes of nonstationarity (Maposa et al., 2016). In non-stationary settings, Hounkpe (2015) provided a statistical model for predicting flood probability, the model was used to five gauging stations in the Queme River Basin in Benin Republic, West Africa and it fits the annual maximum discharge using a time-dependent and covariate-dependent generalized extreme value (GEV) distribution.

In the above works, the GEV distribution was considered to be a good model for non-stationary modeling with different approaches. Nagesh and Laxmi.(2021): identified GEV distribution as a best flood frequency distribution for 12 sites in MRB under the stationary conditions. In this study, extension of Nagesh and Laxmi.(2021) work, by utilizing an extended GEV model q-GEV distribution for the first time as an alternate distribution for modeling the non-stationary situation in MRB. Nagesh and Laxmi.(2023): Estimating the Extreme Flood Height Quantiles through Bayesian Approach using q-GEV distribution in MRB.

In practice, the GEV model may be insufficient, but its generalizations should provide more modelling flexibility. Provost et al. (2018) introduced the extended model q-GEV distribution, which is a q analogue of the GEV, where q is an extra parameter that allows for greater flexibility in modeling extreme events than the GEV distribution. In previous work, the q-GEV distribution was shown that better flood frequency model than GEV distribution for twelve MRB sites using the block maxima technique with stationary assumption. For the same sites, our objective in this study is to model maximum flood height using time dependent q-GEV distribution under non-stationary assumption that is when covariates are present. This study involves a non-stationary q-GEV distribution as time dependent, expecting the variation linearly or nonlinearly over time in location and scale parameters while shape parameter is unchanged with the time. A statistical modeling approach is advocated by considering maximum likelihood estimation (MLE) in the presence of covariates like trends and cycles (Katz et al., 2002). Further related information can be found in the literature (Hesarkazzazi et al., 2021; Meis et al., 2021; Sen, 2021).

As far as we know no similar work has been done in earlier studies on extreme value theory in a changing climate for the Mahanadi River Basin.

## **2. Materials and methods**

This section explains how the data was analyzed. The extended time homogeneous q-GEV distribution was used to investigate linear and quadratic trend models.

### **The data**

The original form of twelve sites flood height data recorded thrice a day was provided by the Central Water Commission (CWC), Bhubaneswar, which is the competent authority for water resources management in the river basin. Using successive steps in each hydrological year, the annual maximum series was obtained. Ferreira and De Haan (2015) provide detailed descriptions of the block maxima probability theory as well as practical considerations for choosing block maxima rather than peak over threshold, Dombry (2013) demonstrated that when utilizing the block maxima technique ML estimators are consistent. Block maxima and Peak over Threshold are two important methodologies used in extreme value theory for flood frequency analysis. When the sample size is large, the block maxima approach is utilized, in which each hydrological year is considered as its own block (Ferreira and De Haan, 2015).

### **Non-stationary extreme value models**

The goal of the study was to model the behavior of annual maximum flood heights in the presence of covariates, in order to see if non-stationary models fit the data better than stationary (time independent) models. Trends, cycles, and physical factors are examples of covariates, according to Katz et al. (2002). Trends are considered as covariates in particular, and the method of time varying moments is the most commonly employed approach to non-stationary flood frequency analysis.

The models  $M_0$ ,  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  were considered to compare the stationary (Time independent) and non-stationary (Time dependent) models, where

- $M_0$  - q-GEV time independent (stationary) model,
- $M_1$  - q-GEV model with linear trend in location and scale parameter,
- $M_2$  - q-GEV model with linear trend in location parameter,
- $M_3$  - q-GEV model with linear trend in scale parameter and
- $M_4$  - q-GEV model with non-linear trend in location parameter.

In previous work, the best flood frequency model was shown to be the q-GEV (time independent) distribution at twelve MRB sites (Nagesh and Laxmi, 2022). This study looks at non-stationary (time-dependent) models for the same sites.  $M_0$  is a stationary model that is not affected by time. Non-stationary models  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  have a linear or nonlinear trend in the location or scale parameter, or both.

The distribution function and density function of q-GEV distribution is given by

$$F(x; s, m, \xi, q) = \left[ 1 + q(1 + \xi(xs - m))^{\frac{1}{\xi}} \right]^{-\frac{1}{q}} ; \quad \xi \neq 0, q \neq 0 \quad (1)$$

$$f(x; s, m, \xi, q) = s(1 + \xi(xs - m))^{(-1/\xi)-1} [1 + q(1 + \xi(xs - m))^{\frac{1}{\xi}}]^{(-1/q)-1} ; \quad \xi \neq 0, q \neq 0 \quad (2)$$

The log-likelihood function of q-GEV distribution is given by

$$l(s, m, \xi, q) = n * \log(s) + \left(-\frac{1}{q} - 1\right) \sum_{i=1}^n \log \left[ q(\xi(x_i s - m) + 1)^{-1/\xi} + 1 \right] + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^n \log \xi x_i s - m + 1 \quad (3)$$

where  $m = \mu/\sigma$  and  $s = 1/\sigma$ ,  $\mu$  is location parameter,  $\sigma$  is scale parameter,  $\xi$  and  $q$  are shape parameters.

The distribution function, Probability density function and log-likelihood function of q-GEV distribution without re-parameterization are given by equations (4), (5) and (6) respectively.

$$F(x; \mu, \sigma, \xi, q) = \left\{ 1 + q \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}^{-1/q} ; \quad \xi \neq 0, q \neq 0 \quad (4)$$

$$f(x; \mu, \sigma, \xi, q) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{(-1/\xi)-1} \left\{ 1 + q \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}^{(-1/q)-1} ; \quad \mu \in R, \sigma > 0, \xi \neq 0, q \neq 0 \quad (5)$$

$$l(\mu, \sigma, \xi, q) = n * \log(\sigma) + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^n \log \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right] + \left(-\frac{1}{q} - 1\right) \sum_{i=1}^n \log \left\{ 1 + q \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (6)$$

$M_1$  is non-stationary q-GEV model with linear trend in both location and scale parameter is

$$\mu(t) = \mu_0 + \mu_1 t, \log \sigma(t) = \sigma_0 + \sigma_1 t, \quad \xi(t) = \xi \quad \text{and} \quad q(t) = q$$

The distribution function of  $M_1$  is given by equation (7) while the log-likelihood function of  $M_1$  is given by equation (8).

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + q \left[ 1 + \xi \left( \frac{x - (\mu_0 + \mu_1 t)}{\exp(\sigma_0 + \sigma_1 t)} \right) \right]^{-1/\xi} \right\}^{-1/q}; \quad \xi \neq 0, q \neq 0 \quad (7)$$

$$l(\mu(t), \sigma(t), \xi(t), q(t)) = n * \log \left( -\frac{1}{\xi} - 1 \right) \sum_{i=1}^n \log \left[ 1 + \xi \left( \frac{x_i - (\mu_0 + \mu_1 t)}{\exp(\sigma_0 + \sigma_1 t)} \right) \right] + \left( -\frac{1}{q} - 1 \right) \sum_{i=1}^n \log \left\{ 1 + q \left[ \xi \left( \frac{x_i - (\mu_0 + \mu_1 t)}{\exp(\sigma_0 + \sigma_1 t)} \right) \right]^{-1/\xi} \right\} \quad (8)$$

where  $t$  is time (in years). The log-likelihood function of  $M_1$  is given by equation (8), and the set of values is estimated by maximization of the log-likelihood function. The Newton Raphson method is used to solve the log-likelihood function equations.

$M_2$  is non-stationary (time dependent) q-GEV model with linear trend in the location parameter.

$$\mu(t) = \mu_0 + \mu_1 t, \sigma(t) = \sigma, \xi(t) = \xi \quad \text{and} \quad q(t) = q$$

Cumulative distribution function and log-likelihood function of  $M_2$  is given by equation (9) and equation (10) respectively

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + q \left[ 1 + \xi \left( \frac{x - (\mu_0 + \mu_1 t)}{\sigma} \right) \right]^{-1/\xi} \right\}^{-1/q}; \quad \xi \neq 0, q \neq 0 \quad (9)$$

$$l(\mu(t), \sigma(t), \xi(t), q(t)) = n * \log(\sigma) + \left( -\frac{1}{\xi} - 1 \right) \sum_{i=1}^n \log \left[ 1 + \xi \left( \frac{x_i - (\mu_0 + \mu_1 t)}{\sigma} \right) \right] + \left( -\frac{1}{q} - 1 \right) \sum_{i=1}^n \log \left\{ 1 + q \left[ \xi \left( \frac{x_i - (\mu_0 + \mu_1 t)}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (10)$$

$M_3$  is non-stationary q-GEV model with linear trend in the scale parameter.

$$\log \sigma(t) = \sigma_0 + \sigma_1 t, \mu(t) = \mu, \xi(t) = \xi \quad \text{and} \quad q(t) = q$$

Distribution function and log-likelihood function of  $M_3$  model are obtained by replacing above quantities in equations (4) and (6) respectively.

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + q \left[ 1 + \xi \left( \frac{x - \mu}{\exp(\sigma_0 + \sigma_1 t)} \right) \right]^{-1/\xi} \right\}^{-1/q}; \quad \xi \neq 0, q \neq 0 \quad (11)$$

$$l(\mu(t), \sigma(t), \xi(t), q(t)) = n * \ln(\sigma_0 + \sigma_1 t) + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^n \log \left[ 1 + \xi \left( \frac{x_i - \mu}{\exp\left(\frac{\sigma_0 + \sigma_1 t}{\xi}\right)} \right) \right] + \left(-\frac{1}{q} - 1\right) \sum_{i=1}^n \log \left\{ 1 + q \left[ \xi \left( \frac{x_i - \mu}{\exp\left(\frac{\sigma_0 + \sigma_1 t}{\xi}\right)} \right) \right]^{-1/\xi} \right\} \quad (12)$$

$M_4$  is non-stationary q-GEV model with non-linear quadratic trend in the location parameter.

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2, \quad \sigma(t) = \sigma, \quad \xi(t) = \xi \quad \text{and} \quad q(t) = q$$

Similarly by replacing above quantities in equations (4) and (6) we can get distribution and log-likelihood function of  $M_4$  q-GEV Model.

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + q \left[ 1 + \xi \left( \frac{x - (\mu_0 + \mu_1 t + \mu_2 t^2)}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}^{-\frac{1}{q}}; \quad \xi \neq 0, q \neq 0 \quad (13)$$

$$l(\mu(t), \sigma(t), \xi(t), q(t)) = n * \log(\sigma) + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^n \log \left[ 1 + \xi \left( \frac{x_i - (\mu_0 + \mu_1 t + \mu_2 t^2)}{\sigma} \right) \right] + \left(-\frac{1}{q} - 1\right) \sum_{i=1}^n \log \left\{ 1 + q \left[ \xi \left( \frac{x_i - (\mu_0 + \mu_1 t + \mu_2 t^2)}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (14)$$

Using the maximum likelihood estimation technique, parameters of the models  $M_0$ ,  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  are calculated using MLE technique. MLE is used to estimate the parameters of both stationary and non-stationary q-GEV models. Using the MLE methodology in the presence of variables is reliable in both the Block Maxima and Peaks over Threshold procedures (Yilmaz et al., 2014; Coles, 2001).

### Model choice

To compare one model to the other, the MLE of nested models employs a simple procedure known as the Deviance (D) statistic (Yilmaz et al., 2014; Coles, 2001; Smith, 1987). The time-homogeneous GEV model,  $M_0$ , is a subset of the time-dependent models  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  in this research. To check significance of non-stationary models over stationary model D statistic is given by

$$D = 2\{l_i(M_i) - l_0(M_0)\} \quad (15)$$

where  $l_i(M_i)$  is the maximized negative log-likelihood value of  $i^{th}$  model ( $i = 1, 2, 3, 4$ ),  $l_0(M_0)$  is the maximized negative log-likelihood value of time independent (stationary) model. D follows chi-square distribution with k degrees of freedom, where k is difference in number of parameters. If  $D > \chi_{k, \alpha}^2$ , then we reject  $H_0$ . It suggests that  $M_i$  is more significant than  $M_0$ .

### 3. Results and discussion

The parameters of the models  $M_0$ ,  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  were calculated using MLE. Parameter estimates of non-stationary q-GEV model and Deviance statistics calculated for pairs of stationary and non-stationary models for Bamnidhi site are given in Table-1. For Bamnidhi site,

consider models  $M_0$  and  $M_1$ , negative log-likelihood value for  $M_0$  and  $M_1$  is -52.9804 and -51.4985 respectively.

Using equation (15),  $D=2[(-51.4985) - (-52.9804)] = 2.9639$ , which is less than 5.991, i.e.  $\chi^2_{2, 0.05} = 5.991$ . So we can conclude that  $M_0$  is better fit than  $M_1$  for Bannidhi site.

**Table-1: Parameter estimates of annual maximum time heterogeneous q-GEV models for Bannidhi site**

Model	$\mu_0$	$\mu_1$	$\mu_2$	$\sigma_0$	$\sigma_1$	$\xi$	q	ll	D
$M_0$	4.6943	0	0	0.8055	0	-0.5750	1.7019	-52.9804	
$M_1$	1.4083	1.408	0	0.9667	0.4028	-0.6900	1.5317	-51.4985	2.9639
$M_2$	4.2001	0.261	0	0.6427	0	-0.4263	1.2799	-51.5775	2.8058
$M_3$	4.8455	0	0	0.7180	-0.0287	-0.7785	1.0460	-51.9625	2.0359
$M_4$	4.6503	1.258	0.217	0.6620	0	-1.1672	2.6899	-52.6737	0.6135

In other words non-stationary model with linear trend in location parameter is insignificant. It clearly shows that the non-stationary model does not give any improvement in fit over the time-homogeneous q-GEV model. Here we can conclude that  $M_0$  is better fit than  $M_1$ . Next consider  $M_0$  and  $M_2$ .

Negative log-likelihood value for  $M_0$  and  $M_2$ , is -52.9804 and -51.5775 respectively.  $D = 2[(-51.5775) - (-52.9804)] = 2.8058$ , which is less than 3.8414.  $\chi^2_{1,0.05} = 3.8414$ .

Similarly the negative log-likelihood value and D statistics values for other pairs of models ( $M_0$ ,  $M_3$ ) and ( $M_0$ ,  $M_4$ ) are given in Table-1. The D statistic values are less than critical values (i.e.  $2.0359 < 3.8414$  and  $0.6135 < 5.991$ ) for the models respectively, p-values when  $\mu=0$  and  $\sigma=0$  are not less than 0.05 for the above models. The models ( $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$ ) do not provide any improvement in fit over the time-homogeneous q-GEV model at Bannidhi site. Therefore stationary model for Bannidhi site is given by (using equation 4).

$$F(x; \mu, \sigma, \xi, q) = \left\{ 1 + 1.7019 \left[ 1 - 0.575 \left( \frac{x_i - 4.6943}{0.8055} \right) \right]^{1/0.575} \right\}^{-1/1.7019}$$

**Table-2: Parameter estimates of annual maximum time heterogeneous q-GEV models for Dharamjaigarh site**

Model	$\mu_0$	$\mu_1$	$\mu_2$	$\sigma_0$	$\sigma_1$	$\xi$	q	ll	D
$M_0$	6.0787	0	0	0.770 7	0	0.340 4	0.650 6	- 44.762 2	
$M_1$	6.0180	0.607 9	0	0.616 6	0.462 4	0.310 6	0.942 5	- 43.131 3	3.2618
$M_2$	5.8273	0.056 2	0	0.377 6	0	0.254 2	0.934 9	- 44.264 8	0.9947
$M_3$	6.7685	0	0	0.698 2	- 0.049	0.201 4	0.743 3	- 42.648 2	4.2279
$M_4$	5.5320	0.217 0	0.00 6	3.009 0		0.313 8	0.891 1	- 43.995 8	1.5327

Consider the models ( $M_0, M_1$ ), D statistic is 3.2618 which is less than  $\chi_{2,0.05}^2 = 5.991$

Linear trend in location parameter is not significant at 5% level of significance. D statistic value for the models ( $M_0, M_2$ ) is 0.9947, which is less than 3.8414(from Table-2). D statistic value for the models ( $M_0, M_3$ ) is 4.2279, which is greater than 3.8414.

The likelihood ratio test for  $\sigma_1 = 0$  has p-value 0.0148 (indicates that  $M_3$  is significant). It is worthwhile to consider time-heterogeneous (non-stationary) model  $M_3$  that is linear trend in scale parameter at Dharamjaigarh. So we conclude that non-stationary model ( $M_3$ ) outperform than stationary model.

Therefore non-stationary model with linear trend in scale parameter for Dharamjaigarh is given by (using equation 11)

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 0.7433 \left[ 1 + 0.2014 \left( \frac{x_i - 6.7685}{\exp(0.6982 - 0.0489 t)} \right) \right]^{-1/0.2014} \right\}^{-1/0.7433}$$

Parameter estimates of stationary and non-stationary q-GEV models for remaining ten sites are given in Table-3.

**Table-3: Parameter estimates of time heterogeneous q-GEV models for ten sites**

Site name	Model	$\mu_0$	$\mu_1$	$\mu_2$	$\sigma_0$	$\sigma_1$	$\xi$	$q$
Kesinga	$M_0$	10.0670	0	0	1.8674	0	-0.6814	1.2930
	$M_1$	10.9203	1.0109	0	1.8639	0.1868	-0.6156	1.1996
	$M_2$	11.4792	1.0358	0	2.6923	0	-0.2112	2.4827
	$M_3$	10.2259	0	0	1.954	0.6547	-0.6952	2.3249
	$M_4$	10.0927	-4.2634	0.2132	1.9666	0	-0.3211	1.3031
Kotni	$M_0$	9.5410	0	0	1.7674	0	-0.5996	1.2471
	$M_1$	9.1822	0.9634	0	1.7580	0.1770	-0.6254	1.1993
	$M_2$	9.1917	1.0535	0	1.8565	0	-0.3397	1.5492
	$M_3$	8.5328	0	0	0.3842	-0.031	-0.7229	1.1002
	$M_4$	9.9001	-0.7310	0.0416	0.5767	0	-0.4306	2.0120
Manendragarh	$M_0$	4.3112	0	0	0.5755	0	0.1531	0.2173
	$M_1$	4.1216	0.0431	0	0.5180	0.4604	0.1791	0.1337
	$M_2$	4.3063	0.4574	0	0.5121	0	0.0772	0.2843
	$M_3$	4.8871	0	0	0.4896	-0.048	0.0627	0.1491
	$M_4$	4.5419	0.0149	0.0043	0.5528	0	0.1738	0.2498
Mohana	$M_0$	3.5996	0	0	0.6233	0	0.4391	1.2296
	$M_1$	3.4053	0.2746	0	0.9942	0.1221	0.3932	2.1456
	$M_2$	3.6147	0.1662	0	0.6176	0	0.2903	2.9925
	$M_3$	3.2420	0	0	0.3677	-0.005	0.3524	1.1606
	$M_4$	3.1959	-0.1583	0.0100	0.6223	0	0.4899	2.6569
Pathardhi	$M_0$	6.3879	0	0	1.4162	0	-0.4998	1.8411
	$M_1$	6.4684	0.6390	0	1.3623	0.1432	-0.5901	1.8992
	$M_2$	6.2766	0.7046	0	0.5444	0	-0.6028	1.9535

	M <sub>3</sub>	6.4457	0	0	4.5827	-0.657	-0.5179	0.9781
	M <sub>4</sub>	3.2121	0.8032	0.0623	0.2218	0	-0.7954	1.8510
Rajim	M <sub>0</sub>	6.2691	0	0	1.1429	0	-0.4184	0.7030
	M <sub>1</sub>	6.4530	0.6197	0	1.1333	0.1087	-0.7899	0.8978
	M <sub>2</sub>	7.0967	0.8451	0	1.0812	0	-0.6399	0.9018
	M <sub>3</sub>	6.9983	0	0	0.8316	0.8161	-0.5623	0.1563
	M <sub>4</sub>	7.9598	-0.2092	0.0634	0.0966	0	-0.7796	0.9472
Seorinarayan	M <sub>0</sub>	14.1171	0	0	2.0398	0	-1.2899	1.1590
	M <sub>1</sub>	14.0027	1.4109	0	2.0608	0.2040	-1.1201	1.1992
	M <sub>2</sub>	14.2159	1.458	0	1.9547	0	-1.654	1.2564
	M <sub>3</sub>	13.9995	0	0	2.0956	0.2354	-1.441	1.9547
	M <sub>4</sub>	14.2555	1.8647	0.0947	2.1587	0	-1.569	2.458
Simga	M <sub>0</sub>	11.0891	0	0	1.9198	0	-0.6793	1.2716
	M <sub>1</sub>	11.2147	1.1083	0	1.7283	0.1920	-0.7590	1.3015
	M <sub>2</sub>	10.4293	0.1565	0	0.9750	0	-0.2570	1.8502
	M <sub>3</sub>	12.2545	0	0	0.1885	-0.038	-0.2824	1.3522
	M <sub>4</sub>	11.9654	-4.1557	0.0978	0.6826	0	-0.1515	1.1974
Sundargarh	M <sub>0</sub>	7.3615	0	0	0.7668	0	-0.2991	0.5387
	M <sub>1</sub>	7.1661	1.0138	0	0.7619	0.0768	-0.2665	0.4889
	M <sub>2</sub>	7.7778	0.9919	0	0.8079	0	-0.2408	0.5498
	M <sub>3</sub>	9.9346	0	0	1.0730	-0.267	-2.1195	0.5479
	M <sub>4</sub>	7.2849	-1.4298	0.0965	0.3273	0	-0.2099	0.7350
Alipingal	M <sub>0</sub>	11.4711	0	0	2.2865	0	-1.2679	3.7158
	M <sub>1</sub>	11.3595	1.1470	0	2.0579	0.2286	-1.1936	3.5176
	M <sub>2</sub>	11.6941	0.5456	0	0.8858	0	-0.0250	3.9536
	M <sub>3</sub>	11.254	0	0	0.5648	0.198	-1.954	3.5578
	M <sub>4</sub>	11.7532	3.6418	0.0317	0.5422	0	-1.1161	3.8538

Table-4 gives Negative log-likelihood (NLL), D statistic values and  $\chi^2$  critical value of stationary and non-stationary time series models for remaining ten sites.

**Table-4: Annual maximum time heterogeneous q-GEV models for ten sites**

Site name	Model	NLL	D	$\chi^2$ Critical Value
Kesinga	M <sub>0</sub>	-83.9958		
	M <sub>1</sub>	-83.1	1.7916	5.991
	M <sub>2</sub>	-83.3512	1.2892	3.8414
	M <sub>3</sub>	-81.5821	4.8274	3.8414
	M <sub>4</sub>	-83.2325	1.5266	5.991
Kotni	M <sub>0</sub>	-82.4308		
	M <sub>1</sub>	-82.1537	0.5542	5.991
	M <sub>2</sub>	-82.3325	0.1966	3.8414
	M <sub>3</sub>	-80.2243	4.413	3.8414
	M <sub>4</sub>	-81.9637	0.9342	5.991
Manendragarh	M <sub>0</sub>	-44.2274		
	M <sub>1</sub>	-43.5480	1.35876	5.991
	M <sub>2</sub>	-43.9629	0.52896	3.8414

	M <sub>3</sub>	-43.649	1.15676	3.8414
	M <sub>4</sub>	-43.3618	1.73116	5.991
Mohana	M <sub>0</sub>	-39.6943		
	M <sub>1</sub>	-39.547	0.2946	5.991
	M <sub>2</sub>	-38.9243	1.54	3.8414
	M <sub>3</sub>	-39.254	0.8806	3.8414
	M <sub>4</sub>	-39.1657	1.0572	5.991
Pathardhi	M <sub>0</sub>	-51.1474		
	M <sub>1</sub>	-50.945	0.4047	5.991
	M <sub>2</sub>	-48.3295	5.6357	3.8414
	M <sub>3</sub>	-50.8495	0.5957	3.8414
	M <sub>4</sub>	-50.5798	1.1351	5.991
Rajim	M <sub>0</sub>	-79.2463		
	M <sub>1</sub>	-78.5139	1.4648	5.991
	M <sub>2</sub>	-79.1537	0.1852	3.8414
	M <sub>3</sub>	-78.9381	0.6164	3.8414
	M <sub>4</sub>	-78.349	1.7946	5.991
Seorinarayan	M <sub>0</sub>	-66.772		
	M <sub>1</sub>	-66.2973	0.9494	5.991
	M <sub>2</sub>	-66.4967	0.5506	3.8414
	M <sub>3</sub>	-63.9427	5.6586	3.8414
	M <sub>4</sub>	-65.5468	2.4504	5.991
Simga	M <sub>0</sub>	-100.255		
	M <sub>1</sub>	-99.5826	1.345	5.991
	M <sub>2</sub>	-98.2431	4.024	3.8414
	M <sub>3</sub>	-99.5488	1.4126	3.8414
	M <sub>4</sub>	-99.2849	1.9404	5.991
Sundargarh	M <sub>0</sub>	-54.8368		
	M <sub>1</sub>	-53.9144	1.8448	5.991
	M <sub>2</sub>	-54.015	1.6436	3.8414
	M <sub>3</sub>	-54.3521	0.9694	3.8414
	M <sub>4</sub>	-53.9986	1.6764	5.991
Alipingal	M <sub>0</sub>	-66.6641		
	M <sub>1</sub>	-66.315	0.6981	5.991
	M <sub>2</sub>	-66.4983	0.3315	3.8414
	M <sub>3</sub>	-66.3219	0.68424	3.8414
	M <sub>4</sub>	-66.1999	0.92832	5.991

At Kesinga site, D statistic of models (M<sub>0</sub>, M<sub>3</sub>) is 4.8274 which is greater than tabulated value 3.8414, it shows that model M<sub>3</sub> is better than stationary model. The D statistic values for all other models except model (M<sub>0</sub>, M<sub>3</sub>) are less than tabulated value. Therefore at Kesinga non-stationary q-GEV model with linear trend in scale parameter M<sub>3</sub> gives improvisation over stationary model M<sub>0</sub>.

So non-stationary model with linear trend in scale parameter for Kesinga site is given by

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 0.7433 \left[ 1 - 0.6952 \left( \frac{x_i - 10.2259}{\exp(1.954 - 0.6547t)} \right) \right]^{1/0.6952} \right\}^{-1/2.3249}$$

Similarly from Table-4, D statistic value for the pair of model  $M_0$  and  $M_3$  for Kotni and Seorinarayan is 4.413 and 5.6586 respectively, which are greater than tabulated value 3.8414 (i.e.  $D > \chi_{1,0.05}^2$ ). Hence we can conclude that model  $M_3$  is significant than  $M_0$  for Kotni site and Seorinarayan site.

So non-stationary model with linear trend in scale parameter for Kotni site is given by

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 1.1002 \left[ 1 - 0.7229 \left( \frac{x_i - 8.5328}{\exp(0.3842 - 0.031t)} \right) \right]^{1/0.7229} \right\}^{-1/1.1002}$$

Non-stationary model with linear trend in scale parameter for Seorinarayan site is given by

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 1.9547 \left[ 1 - 1.441 \left( \frac{x_i - 13.9995}{\exp(2.0956 + 0.2354t)} \right) \right]^{1/1.441} \right\}^{-1/1.9547}$$

Next, the D statistic values for the pair of models ( $M_0, M_2$ ) of Pathardhi and Simga sites are 5.6357 and 4.024 respectively. D statistic values are greater than 3.8414 (i.e.  $D > \chi_{1,0.05}^2$ ). Hence we can conclude that non-stationary q-GEV model with linear trend in location parameter is better model than stationary model at these sites.

Non-stationary model with linear trend in location parameter for Pathardhi is given by (using equation 9)

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 1.9535 \left[ 1 - 0.6028 \left( \frac{x - (6.2766 + 0.7046t)}{0.5444} \right) \right]^{1/0.6028} \right\}^{-1/1.9535}$$

Non-stationary model with linear trend in location parameter for Simga is given by (using equation 9)

$$F(x; \mu(t), \sigma(t), \xi(t), q(t)) = \left\{ 1 + 1.8502 \left[ 1 - 0.257 \left( \frac{x - (10.4293 + 0.1565t)}{0.9750} \right) \right]^{1/0.257} \right\}^{-1/1.8502}$$

D statistics value for all the pairs of models i.e. ( $M_0, M_1$ ), ( $M_0, M_2$ ) and ( $M_0, M_3$ ), and ( $M_0, M_4$ ) are less than tabulated values (5.991, 3.8414, 3.8414 and 5.991 respectively) for Manendragarh, Mohana, Rajim, Sundaragrh and Alipingal. Therefore we can conclude that at these sites stationary q-GEV model is better than non-stationary models.

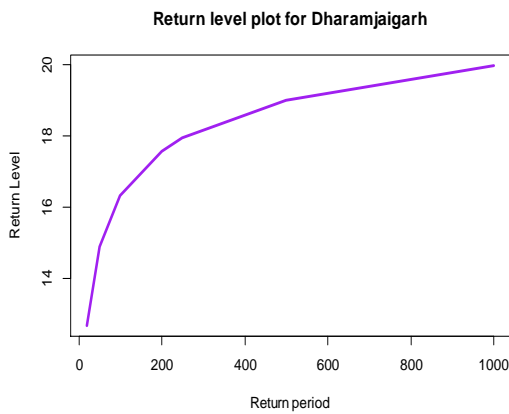
Stationary q-GEV model for Manendragarh is given by

$$F(x; \mu, \sigma, \xi, q) = \left\{ 1 + 0.2173 \left[ 1 + 0.1531 \left( \frac{x - 4.3112}{0.5755} \right)^{-1/0.1531} \right]^{-1/0.2173} \right\}$$

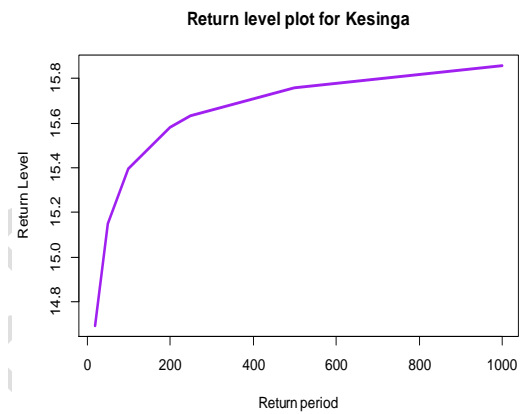
Similarly Stationary q-GEV model for Mohana, Rajim, Sundaragrh and Alipingal are obtained by putting values of  $\mu, \sigma, \xi$  and  $q$  in the equation (4) from Table-3.

D statistic value for  $M_0$  and  $M_4$  values of all the sites are less than 5.991, which indicates that  $M_4$  is nowhere best model among twelve sites. It is better to ignore  $M_4$  at all the sites. Fig. 1 gives return level plot for non-stationary model (i.e. model with linear trend in scale parameter) of Dharamjaigarh, which have higher return levels as compared to stationary model for a different return period. Same nature is found in the other non-stationary model return level plot. The return flood levels at different sites using non-stationary models are higher as compare to that of stationary models.

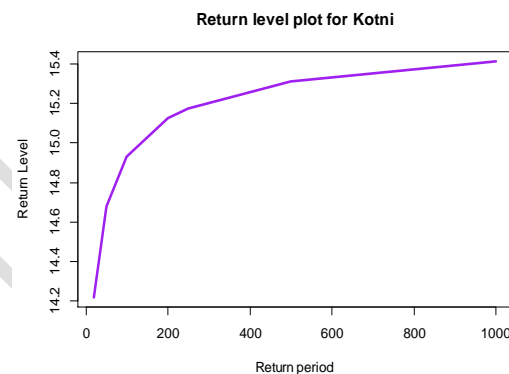
Fig. 1 – Fig. 6 depicts Return level plots for the sites of MRB where non-stationary q-GEV model has improvisation over stationary model.



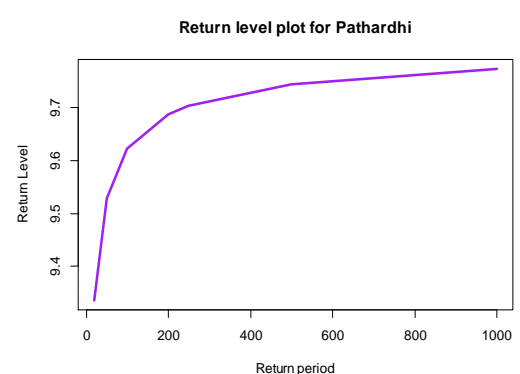
**Fig. 1: Return level plot for Dharamjaigarh site**



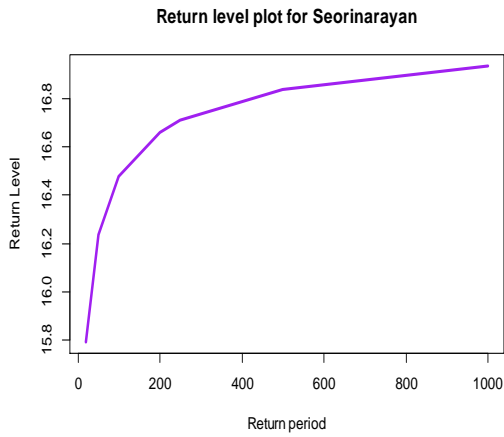
**Fig. 2: Return level plot for Kesinga site**



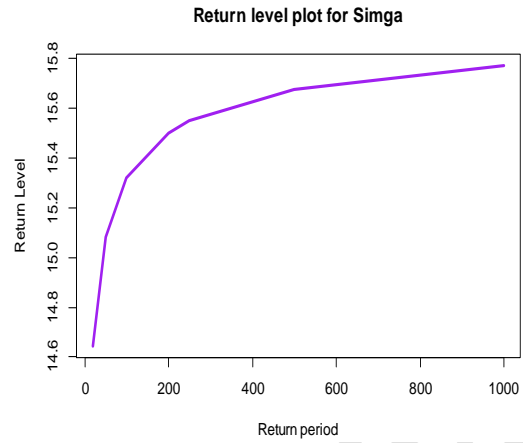
**Fig. 3: Return level plot for Kotni site**



**Fig. 4: Return level plot for Pathardhi site**



**Fig. 5: Return level plot for Seorinarayan site**



**Fig. 6: Return level plot for Simga site**

#### 4. Conclusions

The study looked at the use of extreme value theory in a changing climate for the Mahanadi River Basin in India. The study looked at twelve hydrometric stations that represented twelve different sites along the MRB. The parameters of the q-GEV distribution were determined using the maximum likelihood estimation approach when a long-term trend covariate was present. The study revealed the importance of incorporating non-stationary linear and nonlinear trend models when using extreme value theory in a changing climate, as these models provide a significantly better fit than time-homogeneous models. This improvement in fitness is crucial for the government in planning and policy making.

The prevailing models at the twelve sites were successfully identified, six sites have a time-homogeneous GEV model, four sites have a prevailing time-heterogeneous GEV model with a dominant linear trend in the scale parameter, and two sites have a prevailing time-heterogeneous GEV model with a dominating linear trend in the scale parameter, according to this study.

The results of the study demonstrated that the time independent q-GEV model ( $M_0$ ) is much better fit than time heterogeneous q-GEV models at six sites: Alipingal, Bamnidhi, Manendragarh, Mohana, Rajim, and Sundargarh.

At sites Dharamjaigarh, Kesinga, Kotni and Seorinarayan, the time heterogeneous q-GEV model  $M_3$  explains significant variability than  $M_0$ . Non-stationary model with linear trend in scale parameter outperformed than stationary models. At sites Pathardhi and Simga, time heterogeneous q-GEV model  $M_2$  suits well as compare to  $M_0$ . Non-stationary model with linear trend in location parameter,  $M_2$ , is significant as compared to the Stationary model  $M_0$ .

Non-stationary model with Non-linear trend in location parameter does not fit well over stationary models at any site, which suggests that better to drop out this model consideration for application in the MRB.

#### References:

1. Black K.P., Baba M., Mathew J., Chandra, S., Singh S.S., Shankar R., Kurian N.P., Ulrich P., Narayan B., Stanley D.O., Parsons, S., and Ray, G. (2018). Climate change adaptation guidelines for coastal protection and management in India. prepared by FCGANZDEC (New Zealand) for the Global Environment Facility and Asian Development Bank, Volumes, 1.
2. Coles, S. (2001). Classical extreme value theory and models. In *An introduction to statistical modeling of extreme values* (pp. 45-73). Springer, London.
3. Dombry, C. (2013). Maximum likelihood estimators for the extreme value index based on the block maxima method. *arXiv preprint arXiv:1301.5611*.
4. Ferreira, A., & De Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of statistics*, 43(1), 276-298.
5. Gumbel, E. J. (1941). The return period of flood flows. *The annals of mathematical statistics*, 12(2), 163-190.
6. Hesarkazzazi, S., Arabzadeh, R., Hajibabaei, M., Rauch, W., Kjeldsen, T. R., Prosdocimi, I., Castellarin, A., & Sitzenfrei, R. (2021). Stationary vs non-stationary modelling of flood frequency distribution across northwest England. *Hydrological Sciences Journal*, 66(4), 729-744.
7. Hounkpè, J., Diekkrüger, B., Badou, D. F., & Afouda, A. A. (2015). Non-stationary flood frequency analysis in the Ouémé River Basin, Benin Republic. *Hydrology*, 2(4), 210-229.
8. Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8-12), 1287-1304.
9. Maposa, D., Lesaoana, M., & Cochran, J. J. (2016). Modelling non-stationary annual maximum flood heights in the lower Limpopo River basin of Mozambique. *Jambá: Journal of Disaster Risk Studies*, 8(1), 1-9.
10. Meis, M., Llano, M. P., & Rodriguez, D. (2021). Quantifying and modelling the ENSO phenomenon and extreme discharge events relation in the La Plata Basin. *Hydrological Sciences Journal*, 66(1), 75-89.
11. Nagesh S. and Laxmi D.(2021): Identifying suitable probability models for extreme flood heights in Mahanadi River Basin. *Int. J. Agricult.Sci.* Vol.17, No.1, pp. 197-207.
12. Nagesh S. and Laxmi D.(2023): Estimating the Extreme Flood Height Quantiles using Bayesian Approach. *Indian Journal of Ecology*, Vol. 50, No. 1. pp. 210-214.
13. Provost, S. B., Saboor, A., Cordeiro, G. M., & Mansoor, M. (2018). On the q-generalized extreme value distribution. *REVSTAT-Statistical Journal*, 16(1), 45-70.
14. Ray, L. K., Goel, N. K., & Arora, M. (2019). Trend analysis and change point detection of temperature over parts of India. *Theoretical and Applied Climatology*, 138(1), 153-167.
15. Şen, Z. (2021). Conceptual monthly trend polygon methodology and climate change assessments. *Hydrological Sciences Journal*, 66(3), 503-512.
16. Singh, N., & Chinnasamy, P. (2021). Non-stationary flood frequency analysis and attribution of streamflow series: a case study of Periyar River, India. *Hydrological Sciences Journal*, 66(13), 1866-1881.
17. Singh, O., & Kumar, M. (2013). Flood events, fatalities and damages in India from 1978 to 2006. *Natural hazards*, 69(3), 1815-1834.

18. Sivakumar, B. (2016). *Chaos in hydrology: bridging determinism and stochasticity*. Springer.
19. Smith, P. J. (1987). Maximum likelihood estimation of the distribution of length at age. *Biometrics*, 601-615.
20. Yilmaz, A. G., Hossain, I., & Perera, B. J. C. (2014). Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne. *Hydrology and Earth System Sciences*, 18(10), 4065-4076.

UNDER PEER REVIEW