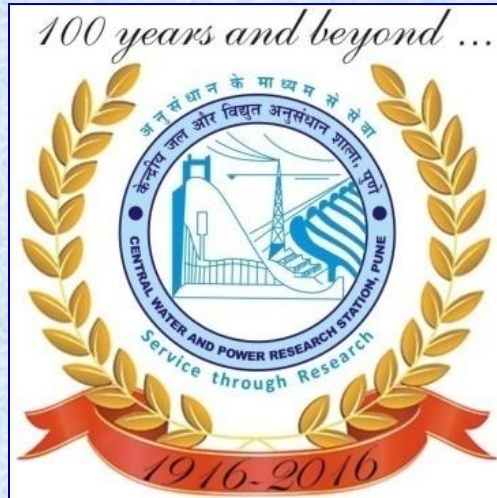


Intercomparison of Normal, Gamma and Extreme Value Families of Probability Distributions for Assessment of Extreme Rainfall



Technical Session
Parallel Session – PW5
Session Theme
Extreme Events and their
Management

N. Vivekanandan*

Central Water and Power Research Station, Pune



**8th International R&D Conference on Global Trends
in Water Resources, Power & RE Sectors
during 8-9 February 2024 at IIT Roorkee**



Outline of the Presentation

- Introduction
- Methodology
 - Extreme Value Analysis (EVA) of Rainfall
 - ❖ Normal family : Normal (N2), Log Normal (LN2)
 - ❖ Gamma family: Gamma (GAM), Pearson Type-3 (PR3)
Log Pearson Type-3 (LP3)
 - ❖ Extreme value family : Extreme Value Type-1 (EV1),
Extreme Value Type-2 (EV2), Generalized Extreme Value (GEV), Generalize Pareto (PRT)
 - ❖ Goodness-of-Fit (GoF) tests viz., Anderson-Darling (A^2) & Kolmogorov-Smirnov (KS) and diagnostic test (D-index).
- Study area and data used
- Results and Discussion
- Recommendation

Introduction

- Assessment of extreme rainfall for a given return period is a pre-requisite for planning, design, management and operation of various hydraulic structures such as dams, bridges, barrages, culverts and storm water drainage systems, etc.
- Depending on the size and design-life of structure, the estimated extreme (i.e., 1-day or 24-hour maximum) rainfall for a given return period is used.
- Extreme Value Analysis (EVA) of rainfall is generally considered as an effective tool to determine the expected rainfall at various chances.
- Such information can be used to prevent floods and droughts, and also considered for planning, design and management of hydraulic structures.

Methodology: Probability Distributions used in EVA

Family	PDF	CDF [F(x) of F(y)]	Quantile estimator (x(T))	
Normal	NOR ($\mu(x)$, $\sigma(x)$)	$F(x) = \Phi\left(\frac{x - \mu(x)}{\sigma(x)}\right)$, $0 < x < \infty$, $\sigma(x) > 0$	$x(T) = \mu(x) + \sigma(x) \Phi^{-1}(1/T)$	Φ^{-1} : Inverse of standard normal distribution
	LN2 ($\mu(y)$, $\sigma(y)$)	$F(y) = \Phi\left(\frac{y - \mu(y)}{\sigma(y)}\right)$, $0 < y < \infty$, $\sigma(y) > 0$	$x(T) = \exp(\mu(y) + \sigma(y) \Phi^{-1}(1/T))$	with $y = \ln(x)$
Gamma	GAM (α , k)	$F(x) = \begin{cases} G\left(k, \frac{x}{\alpha}\right) & , \alpha > 0 \\ 1 - G\left(k, \frac{x}{\alpha}\right) & , \alpha < 0 \end{cases}$	No explicit expression of quantile function is available.	
	PR3 (ξ , α , k)	$F(x) = \begin{cases} G\left(k, \frac{x - \xi}{\alpha}\right) & , \alpha > 0 \\ 1 - G\left(k, \frac{x - \xi}{\alpha}\right) & , \alpha < 0 \end{cases}$	No explicit expression of quantile function is available.	
	LP3 (ξ , α , k)	$F(x) = \begin{cases} G\left(k, \frac{\ln(x) - \xi}{\alpha}\right) & , \alpha > 0 \\ 1 - G\left(k, \frac{\ln(x) - \xi}{\alpha}\right) & , \alpha < 0 \end{cases}$	No explicit expression of quantile function is available.	
Extreme Value	EV1 (ξ , α)	$F(x) = \exp\left(-\exp\left(-\frac{x - \xi}{\alpha}\right)\right)$, $\alpha > 0, -\infty < x < \infty$	$x(T) = \xi - \alpha \ln[-\ln(1 - (1/T))]$	
	EV2 (α , k)	$F(x) = \exp\left(-\left(\frac{x}{\alpha}\right)^{-k}\right)$, $\alpha > 0, k > 0, 0 < x < \infty$	$x(T) = \alpha \left[-\ln(1 - (1/T))\right]^{-1/k}$	$x(T)$: Estimated 24-hour maximum rainfall by probability distribution
	GEV (ξ , α , k)	$F(x) = \exp\left(-\left[1 - \frac{k(x - \xi)}{\alpha}\right]^{-1/k}\right)$, $-\infty \leq x \leq \xi + (\alpha/k)$ for $k > 0$, $\alpha > 0$	$x(T) = \xi + \frac{\alpha}{k} \left(1 - [-\ln(1 - (1/T))]^k\right)$	
	PRT (ξ , α , k)	$F(x) = 1 - \left(1 - \frac{k(x - \xi)}{\alpha}\right)^{1/k}$, $\xi \leq x \leq \xi + (\alpha/k)$ for $k > 0$, $\alpha > 0$	$x(T) = \xi + \frac{\alpha(1 - (1/T)^k)}{k}$	
ξ : Location parameter		α : Scale parameter	k : Scale parameter	
G(...): Incomplete gamma integral		$\mu(x)$: Average of observed data (x)	$\mu(y)$: Average of log-transformed series of observed data	
$F(x)$: Cumulative distribution function of x		$\sigma(x)$: Standard deviation of observed data (x)	$\sigma(y)$: Standard deviation of log-transformed series of observed data	
K(T): Frequency factor for a return period (T) corresponding to probability of exceedance (P) and Coefficient of Skewness (C_s) (i.e., $C_s = 0.0$ for LN2 and C_s of the observed data for PR3)				

Methodology

Goodness-of-Fit and Diagnostic Tests

Goodness-of-Fit (GoF) Tests

$$A^2 = (-N) - (1/N) \sum_{i=1}^N \{(2i-1) \ln(Z(i)) + (2N+1-2i) \ln(1-Z(i))\}$$

$$KS = \text{Max}_{i=1}^N (F_e(x(i)) - F_D(x(i)))$$

Diagnostic Test

$$D\text{-index} = (1/\mu(x)) \sum_{i=1}^6 |x(i) - x^*(i)|$$

Goodness-of-Fit Tests

A² : Anderson-Darling

KS : Kolmogorov-Smirnov

CDF Cumulative Distribution Function

Z(i) $F_e(x(i))$ for $i=1,2,3,\dots,N$ with $x(1)<x(2)<\dots<x(N)$

$F_e(x(i))$ Empirical CDF of $x(i)$ of i^{th} sample [**$F_e(x(i))=r/(N+1)$**]

$F_D(x(i))$ Derived CDF of $x(i)$ of i^{th} sample by probability distribution function

$x(i)$ Observed value of x of i^{th} sample

Diagnostic test

$x(i)$ Observed value of x of i^{th} sample (first six highest values in the observed data series)

$x^*(i)$ Estimated value of i^{th} sample (predicted values by probability distribution function)

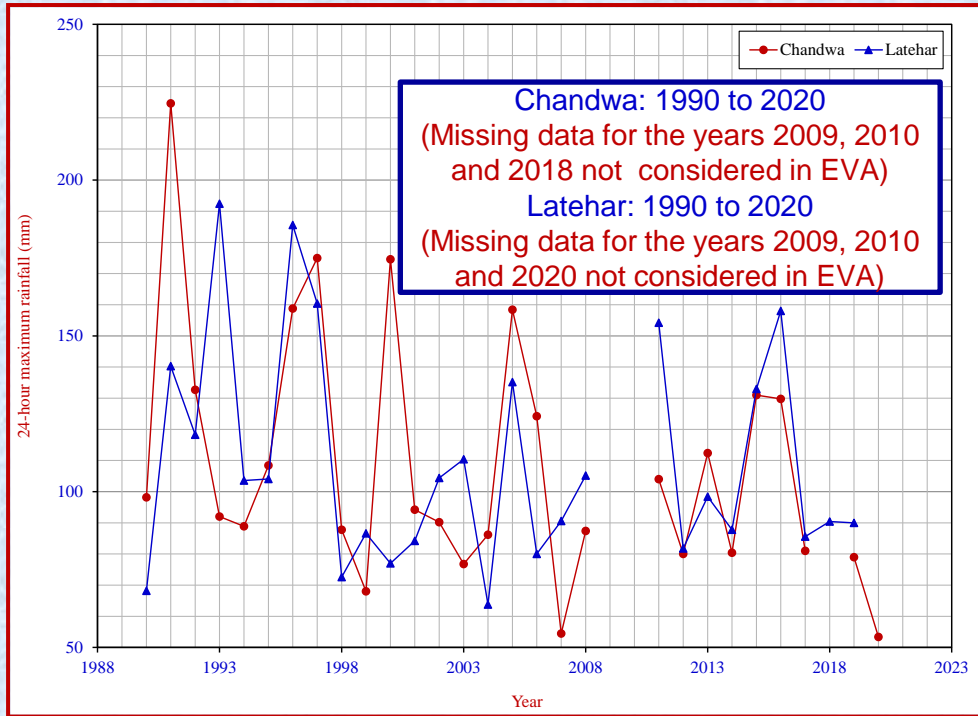
$\mu(x)$ Average of observed values

N Total number of samples

➤ If the computed values of **GoF** tests statistic by the distribution are lesser than its theoretical values at either 5% or 1% level of significance then the distribution is considered to be acceptable for EVA at that level.

➤ Probability distribution with **least D-index** is considered as best for rainfall estimation.

Study Area and Data Used



**Observed 24-hour maximum rainfall
(Chandwa and Latehar)**

- Chandwa rain gauge site is located between latitude $23^{\circ} 40' 12''$ N and longitude $84^{\circ} 44' 56.4''$ E.
- Latehar rain gauge site is located between latitude $23^{\circ} 44' 45.6''$ N and longitude $84^{\circ} 30' 32.4''$ E
- Due to non-availability of daily rainfall data, the hourly rainfall data observed at Chandwa and Latehar sites for the period 1990 to 2020 was used.
- By considering the importance of the hydrological extremes, the rainfall data for the missing years were ignored and also not considered in EVA.

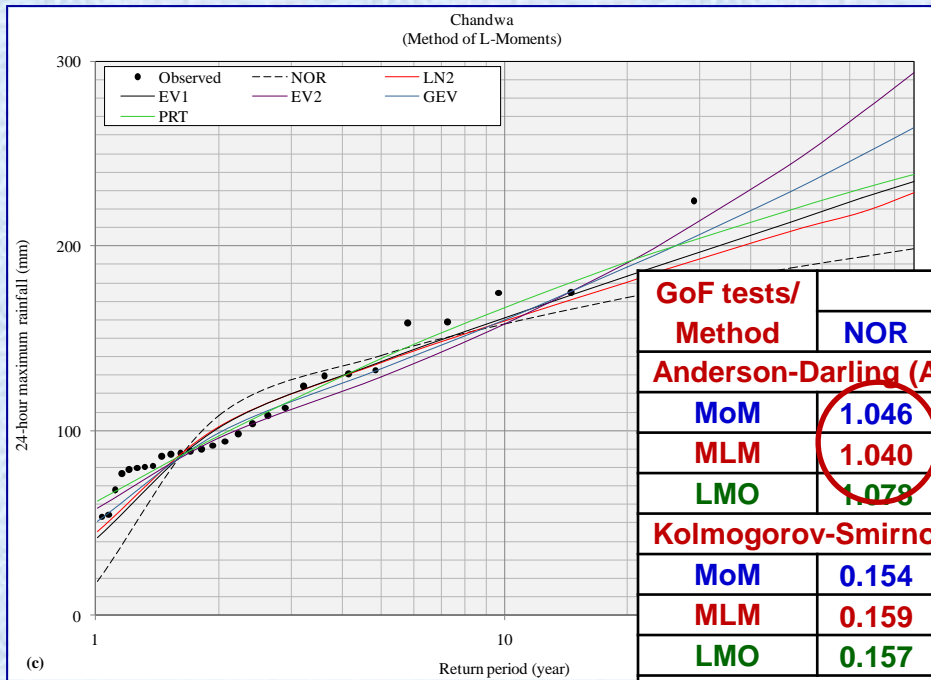
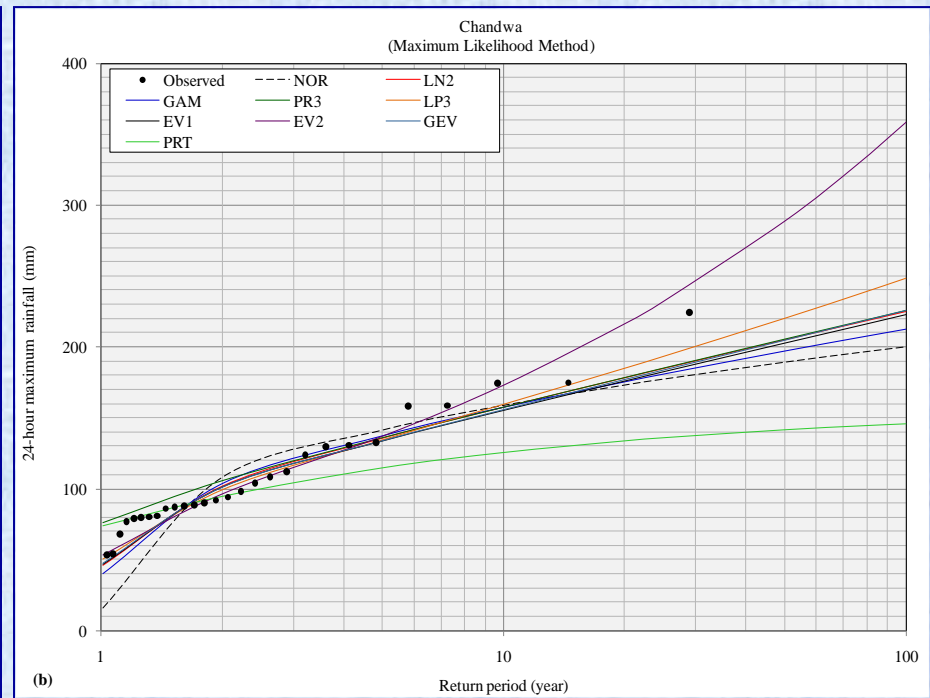
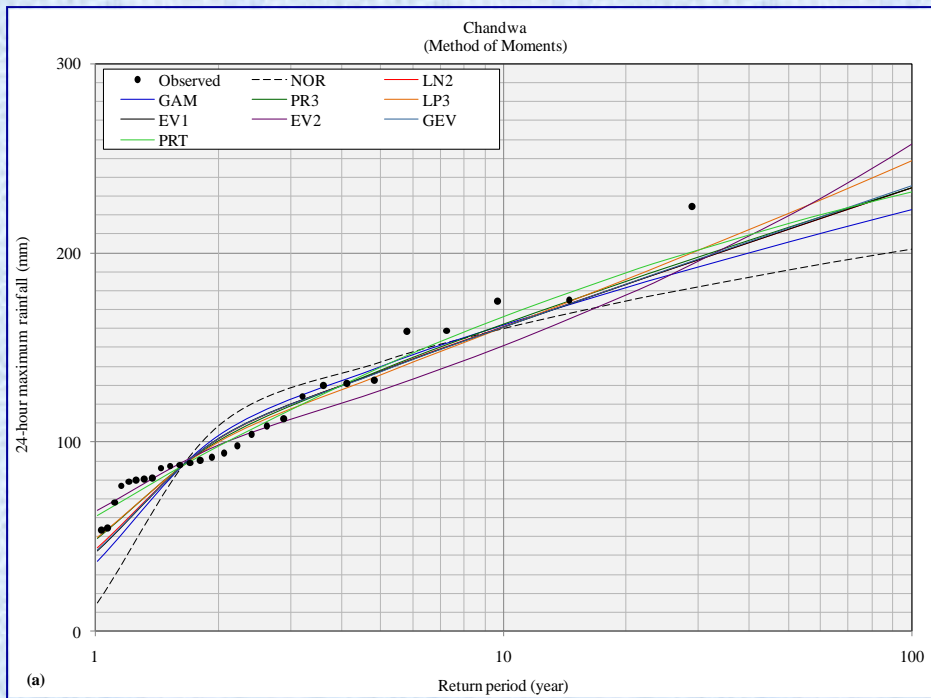
Rain gauge Site	Average (mm)	Std. Deviation (mm)	Coefficient of skewness	Coefficient of kurtosis	Minimum (mm)	Maximum (mm)
Chandwa	108.3 (4.625)	40.2 (0.347)	1.168 (0.337)	1.254 (-0.094)	53.4 (3.978)	224.6 (5.414)
Latehar	109.4 (4.649)	35.1 (0.301)	0.967 (0.491)	0.070 (-0.626)	63.8 (4.156)	192.4 (5.260)

Descriptive statistics of log transformed data of **24-hour maximum rainfall** are given within brackets.

Results and Discussion

Estimated 24HMR for different return periods using Normal, Gamma and Extreme Value families of distributions for Chandwa

Distri- bution	Method	24HMR (in mm) for different return periods								
		1.01	2	5	10	20	25	50	75	100
NOR	MoM	14.6	108.3	142.1	159.8	174.4	178.6	190.8	197.4	201.8
	MLM	16.3	108.3	141.5	158.9	173.2	177.4	189.3	195.8	200.1
	LMO	18.2	108.3	140.8	157.8	171.9	176.0	187.7	194.0	198.3
LN2	MoM	44.0	101.5	137.4	160.9	183.3	190.4	212.3	225.1	234.2
	MLM	46.1	102.0	135.9	157.8	178.6	185.2	205.4	217.1	225.4
	LMO	45.5	102.0	136.6	159.1	180.5	187.3	208.0	218.1	228.7
GAM	MoM	36.9	103.4	139.9	161.9	181.7	187.8	205.8	215.9	222.9
	MLM	40.7	104.1	137.7	157.8	175.7	181.2	197.4	206.5	212.8
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
PR3	MoM	49.1	100.6	137.9	162.2	184.8	191.9	213.3	225.5	234.1
	MLM	76.1	105.7	136.0	157.5	178.4	185.1	205.5	217.3	225.7
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
LP3	MoM	49.6	100.0	135.6	160.8	186.2	194.6	221.1	237.2	248.9
	MLM	50.9	99.7	134.8	159.9	185.3	193.7	220.5	236.8	248.7
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
EV1	MoM	42.2	101.7	137.2	160.7	183.3	190.5	212.5	225.3	234.4
	MLM	46.9	101.4	133.9	155.5	176.2	182.8	203.0	214.7	223.1
	LMO	42.0	101.7	137.3	161.0	183.6	190.8	212.9	225.8	234.9
EV2	MoM	64.0	98.4	127.3	150.9	177.7	187.1	219.5	240.8	257.2
	MLM	53.9	96.8	137.4	173.3	216.5	232.4	288.8	327.7	358.3
	LMO	58.0	95.8	129.3	157.8	190.9	202.8	244.3	272.3	293.9
GEV	MoM	42.9	101.6	137.0	160.6	183.3	190.6	212.9	226.0	235.2
	MLM	47.4	101.1	133.8	155.7	177.0	183.7	204.7	217.0	225.7
	LMO	50.9	98.5	133.6	160.1	188.4	197.9	229.4	249.2	263.9
	MoM	61.1	97.7	139.4	166.1	189.3	196.1	215.3	225.3	231.9



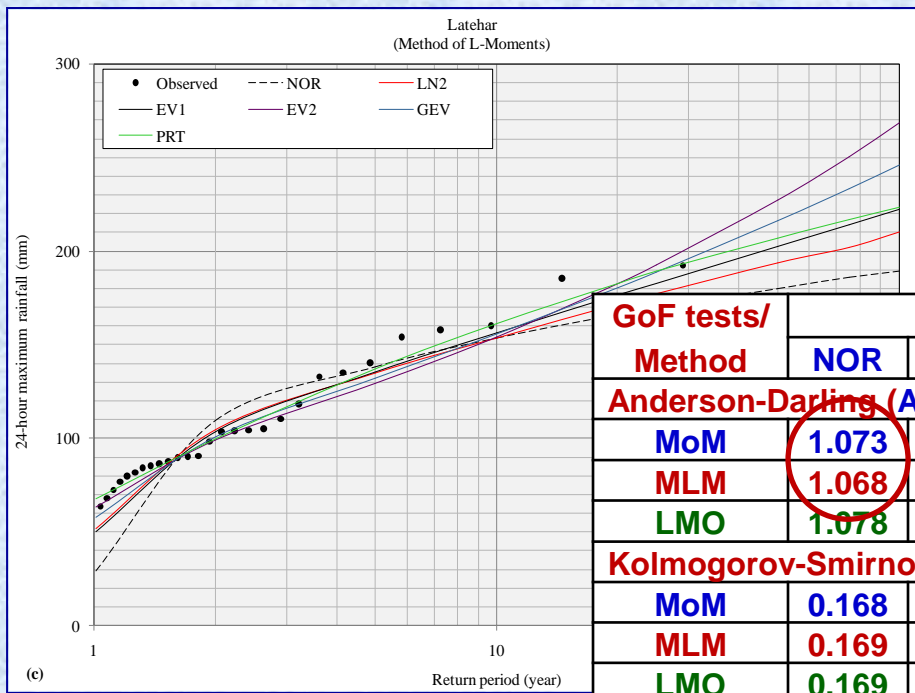
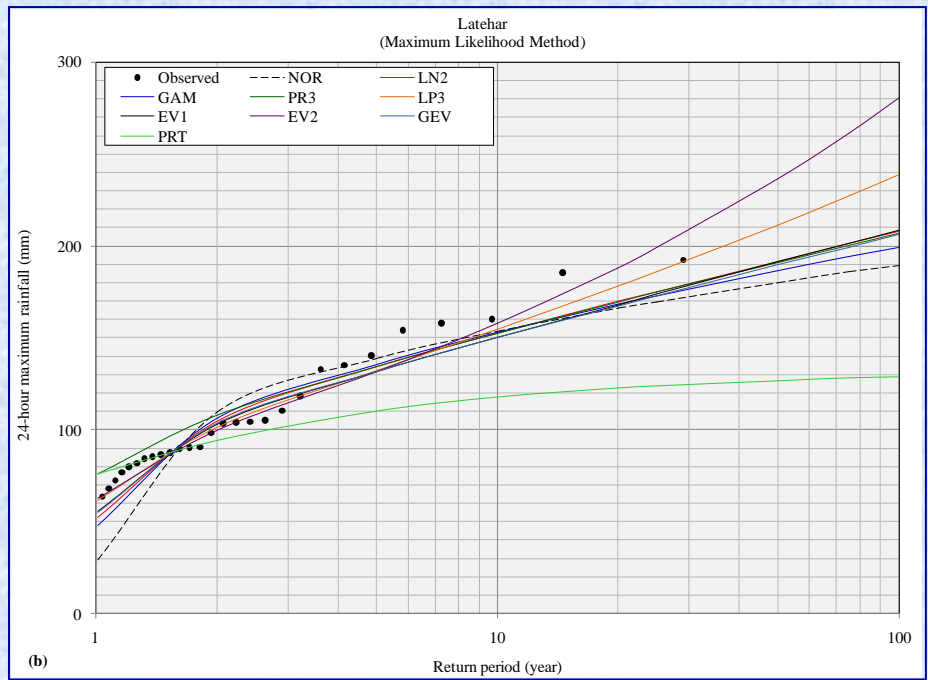
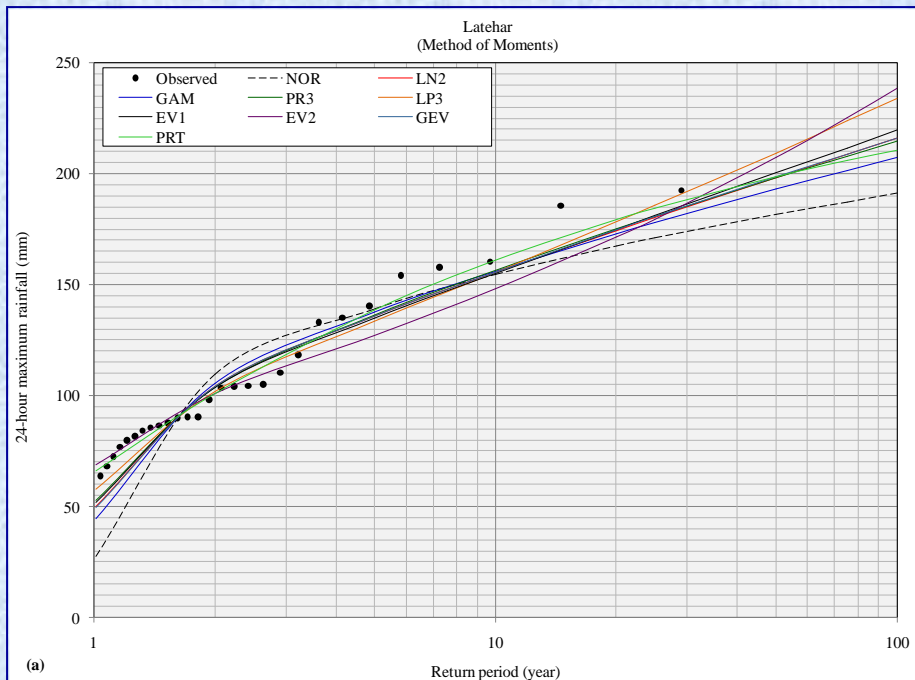
Estimated 24HMR by using Normal, Gamma and Extreme Value families of probability distributions for Chandwa

GoF tests/ Method	Computed values of GoF tests								
	NOR	LN2	GAM	PR3	LP3	EV1	EV2	GEV	PRT
Anderson-Darling (A²)									
MoM	1.046	0.447	0.406	0.356	0.358	0.430	0.559	0.343	7.756
MLM	1.040	0.441	0.403	0.355	0.353	0.410	0.542	0.332	7.652
LMO	1.078	0.452	NF	NF	NF	0.441	0.545	0.345	7.850
Kolmogorov-Smirnov (KS)									
MoM	0.154	0.100	0.113	0.090	0.087	0.102	0.091	0.102	0.178
MLM	0.159	0.100	0.115	0.090	0.091	0.102	0.086	0.088	0.186
LMO	0.157	0.100	0.118	0.090	0.095	0.102	0.081	0.075	0.193

Theoretical values of GoF tests statistic are $A^2_{0.05} = 0.757$ and $KS_{0.05, 28} = 0.250$

Estimated 24HMR for different return periods using Normal, Gamma and Extreme Value families of distributions for Latehar

Distri- bution	Method	24HMR (in mm) for different return periods								
		1.01	2	5	10	20	25	50	75	100
NOR	MoM	27.6	109.4	138.9	154.3	167.1	170.8	181.4	187.1	191.0
	MLM	29.1	109.4	138.4	153.5	166.0	169.7	180.1	185.7	189.5
	LMO	29.1	109.4	138.3	153.5	166.0	169.6	180.1	185.7	189.5
LN2	MoM	50.2	104.1	135.5	155.5	174.2	180.1	198.0	208.4	215.7
	MLM	52.5	104.5	134.0	152.6	169.9	175.2	191.7	201.1	207.7
	LMO	51.9	104.5	134.6	153.6	171.4	176.9	193.8	201.9	210.3
GAM	MoM	44.4	105.6	137.3	156.0	172.7	177.8	192.8	201.2	207.0
	MLM	47.8	106.1	135.5	152.8	168.1	172.7	186.4	194.0	199.3
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
PR3	MoM	52.8	103.8	136.1	156.4	175.0	180.8	198.0	207.8	214.7
	MLM	76.1	108.0	134.5	152.4	169.4	174.8	191.1	200.4	207.0
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
LP3	MoM	57.9	102.0	133.3	155.6	178.1	185.5	209.1	223.5	233.9
	MLM	61.5	101.1	131.8	154.5	178.0	185.9	211.4	227.2	238.8
	LMO	NF	NF	NF	NF	NF	NF	NF	NF	NF
EV1	MoM	51.7	103.6	134.6	155.1	174.8	181.1	200.3	211.5	219.5
	MLM	55.8	103.0	131.3	150.0	167.9	173.6	191.2	201.3	208.6
	LMO	50.3	103.5	135.2	156.3	176.4	182.8	202.5	214.0	222.1
EV2	MoM	68.6	100.9	127.0	147.9	171.3	179.4	207.0	224.9	238.5
	MLM	62.5	99.4	131.3	157.8	188.2	199.1	236.5	261.5	280.7
	LMO	63.2	98.8	129.1	154.1	182.6	192.7	227.4	250.5	268.2
GEV	MoM	49.3	104.2	135.5	155.6	174.5	180.4	198.3	208.6	215.7
	MLM	55.1	103.4	131.6	150.0	167.4	172.9	189.7	199.4	206.2
	LMO	57.7	100.8	132.2	155.6	180.4	188.8	216.2	233.3	245.9
	MoM	66.0	100.9	138.3	160.8	179.2	184.4	198.6	205.8	210.3



Estimated 24HMR by using Normal, Gamma and Extreme Value families of probability distributions for Chandwa

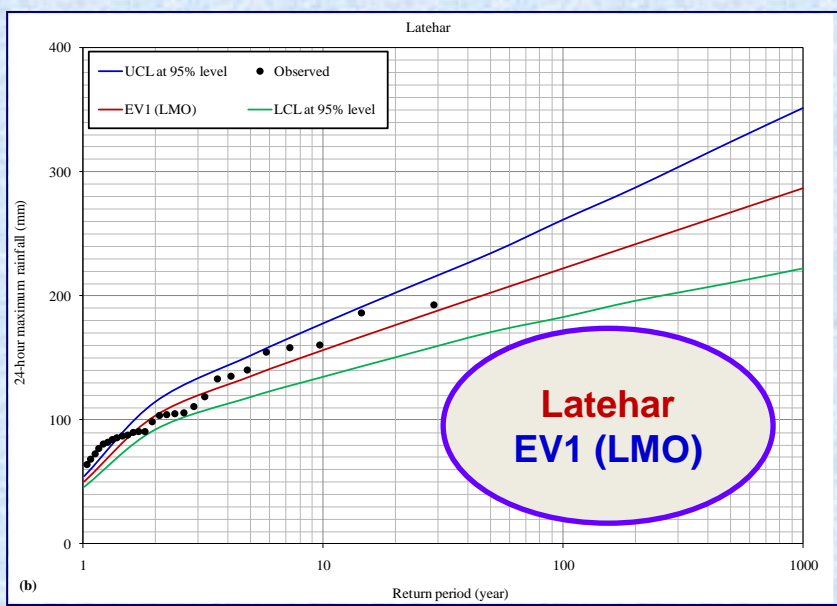
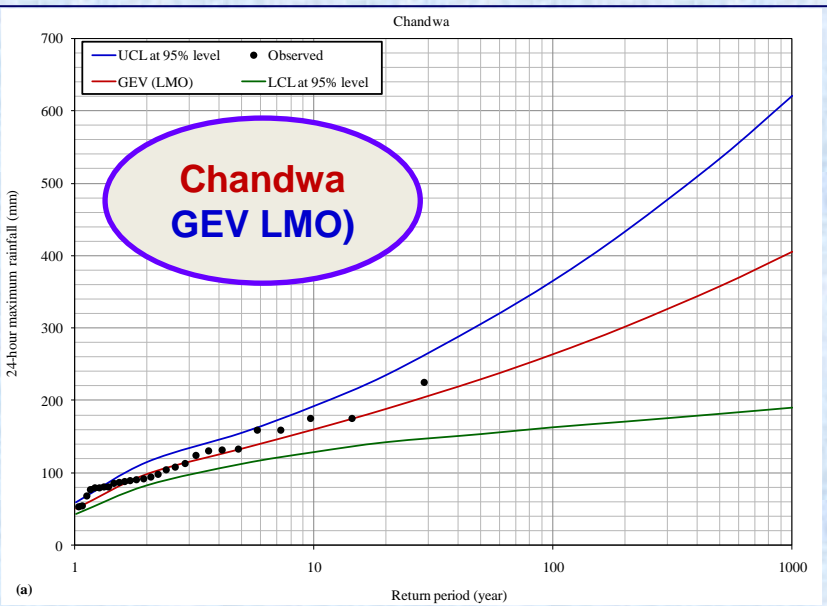
GoF tests/ Method	Computed values of GoF tests								
	NOR	LN2	GAM	PR3	LP3	EV1	EV2	GEV	PRT
Anderson-Darling (A^2)									
MoM	1.073	0.542	0.248	0.248	0.312	0.434	0.273	0.280	4.063
MLM	1.068	0.539	0.245	0.244	0.309	0.426	0.251	0.275	4.053
LMO	1.078	0.545	NF	NF	NF	0.455	0.262	0.305	4.069
Kolmogorov-Smirnov (KS)									
MoM	0.168	0.120	0.126	0.112	0.108	0.120	0.140	0.122	0.076
MLM	0.169	0.121	0.132	0.118	0.121	0.117	0.120	0.110	0.088
LMO	0.169	0.120	NF	NF	NF	0.114	0.098	0.096	0.099

Theoretical values of GoF tests statistic are $A^2 = -0.757$ and $KS = -0.250$

Evaluation of EVA Results using D-index

Computed values of **D-index** using **Normal**, **Gamma** and Extreme Value families of probability distributions for Chandwa and Latehar

Method	D-index values given by								
	NOR	LN2	GAM	PR3	LP3	EV1	EV2	GEV	PRT
Chandwa									
MoM	0.861	0.686	0.694	0.636	0.635	0.689	1.091	0.692	0.560
MLM	0.899	0.826	0.878	0.837	0.670	0.914	0.575	0.896	2.586
LMO	0.942	0.690	NF	NF	NF	0.680	0.690	0.616	0.542
Latehar									
MoM	0.682	0.601	0.585	0.563	0.575	0.615	0.947	0.597	0.373
MLM	0.725	0.760	0.759	0.759	0.624	0.892	0.649	0.897	2.678
LMO	0.726	0.652	NF	NF	NF	0.556	0.732	0.590	0.345



Conclusions

Intercomparison on EVA results of rainfall was performed through **quantitative assessment** (viz., Anderson-Darling, Kolmogorov-Smirnov and D-index) & **quantitative assessment** using fitted curves of estimated rainfall.

- For the return periods from 50-year and above, it was found that the estimated extreme (i.e., 24-hour maximum) rainfall by **EV2** distribution are higher than those values of other distributions considered in the study.
- For **Chandwa** and **Latehar**, it could be seen that the **A² tests results** supported the use of all seven distributions other than **NOR** and **PRT** for **EVA** of rainfall.
- **KS test results** confirmed the applicability of all nine distributions considered in the study for rainfall estimation.
- **Qualitative assessment** of the outcomes was weighed together with **D-index values** and **fitted curves** of the estimated 24-hour maximum rainfall. Accordingly, **GEV (LMO)** is better suited distribution for rainfall estimation for **Chandwa** while **EV1 (LMO)** for **Latehar**.
- From fitted curves of the estimated rainfall, it was observed that the percentage of observed data are lying within confidence limits of the estimated rainfall by the selected probability distribution is about **85%** for **Chandwa** while **75%** for **Latehar**.

Recommendation

Study suggested that the estimated 24-hour maximum rainfall for different return periods obtained from

- GEV (LMO) for Chandwa and
- EV1 (LMO) for Latehar

could be used for planning, design and management of hydraulic structures in the respective sites.

THANK YOU