Hydrologic Design of Soil and Water Conservation Structures Using Probability Analysis and Machine Learning Techniques

ABSTRACT

The soils and water conservation structures are constructed to overcome water scarcity as aresult of interannual rainfall variability and paucity of the perennial source of water. A study was conducted to estimate the design runoff for the efficient hydrologic design of various soils and water conservation structures using Annual One Day Maximum Rainfall (ADMR) for the Saurashtra ~~Region~~ region of Gujarat, India. The design rainfall at various return periods was predicted using three technics *i.e.*~~probability~~ Probability ~~distribution~~ Distribution ~~fitting~~Fitting, Artificial Neural Network (ANN) and Gaussian Process Regression (GPR) for 11 stations. Various goodness of fit tests revealed that ADMR was efficiently predicted by log-logistic (3P) distribution for six stations, generalized extreme value distribution for two stations, ~~and lognormal (and 3P),~~ gamma (3P) and lognormal distribution for one station each. Among ANN and GPR, the performance indicators revealed that GPR has shown a higher capability to predict ADMR as compared to ANN with correlation coefficient ranging from 0.97 to 0.99, mean absolute error from 15 mm to 411 mm and root mean squared error from 40 mm to 494 mm for various stations. The design runoff estimation was demonstrated based on predicted ADMR for various soils and water conservation structures using SCS-Curve Number method for curve number 70 and 85. The results of this study are ~~is~~ useful for researchers, planners and engineers to implement the economical, efficient and safe design of various soils and water conservation structures.

Keywords:*Annual one-day maximum rainfall; Artificial Neural Network; Gaussian Process; Probability distribution; Runoff estimation*

Comment [HO1]: Seven according to your Table2.

1. INTRODUCTION

Formatted: Font: (Default) Times New Roman, 14 pt, Bold, Font color: Black

Formatted: List Paragraph, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

The efficient harvesting and conservation of water is a prominent feature for sustainable development of agriculture, climate change resilience and coping with inter-annual precipitation variability (Bitterman*et al.*, 2016;Gandhi and Patel, 2019). Harvested rainwater is not only useful in dry spells during *the Kharif* seasonbut also for early sowing of *Rabi* crops (Singh *et al.*, 2019). The complete design of soils and water conservation structures can be split into three sections *i.e.* Hydrological Design, Hydraulic Design and Structural Design. The general guidelines are inadequate for executing such designs due to their location-specific nature (Singh and Singh, 2017). The hydrologic design includes design runoff estimation from expected rainfall at a specific recurrence interval. While an over-designed structure results in an uneconomical design along with wastage of area towards construction, underestimated structure possesses a risk of failure, not conserving desired runoff and causing soil erosion in downstream areas (Gundalia, 2016;Otim*et al.*, 2019**)**

The duration of rainstorms in semi-arid regions rarely exceeds one day. Therefore, the annual one-day maximum rainfall (ADMR)is critical for designing soils and water conservation structures. The expected ADMR at various return periods were estimated by fitting various probability distributions e.g. normal, log normal, log pearson type-III, log-logistic, gamma, generalized extreme value, etc. to observed rainfall and location-specific best-fit modelswere selected (Gebremedhin *et al.*, 2017;Dhupal and Swain, 2020). However, Nahvi, *et al.*, (2018) favoredthe requirement of advanced and more reliable computational and simulation methods to model nonlinear and complex phenomena of precipitation. Machine learning techniques like Artificial Neural Networks (ANNs) and Gaussian process regression (GPR) can be prove~~n~~d useful to model complex hydrological processes including rainfall-runoff modeling and precipitation.

Abundant evidence is available for the successful prediction of rainfall/runoffusing ANN. (Darji *et al.*, 2015). Gaussian process regression (GPR) is another non-parametric supervised machine learning method to model complex natural phenomena including rainfall (Ghasemi*et al;* 2021). The GPR outperformed ANN for predicting various metrological and hydrological quantities including rainfall (Chang and Chen, 2018;Ghasemi*et al*., 2021). Mishra and Kushwaha(2019) developed a model to predict rainfall using Gaussian process regression as a classifier with 95.4% accuracy at Raipur, India. Several methods are available to estimate runoff fromrainfall however, United States Soil Conservation Service-Curve Number (SCS-CN) (USDA-SCS, 1986) is widely used due to quick and accurate runoff estimation, simplicity, robustness and acceptability and integration of various parameters in one number i.e. Curve Number. (Singh and Singh,2017;Gandhi and Patel, 2019;Otim *et al.*,2019). The past and recently developed popular hydrological models incorporated SCS-CN method for runoff estimation (Paola De *et al.*, 2013).

The present study was carried out in the Saurashtra region of Gujarat state (India), a semi-arid region with high dependence on rainfall, scare perennial water resources with peculiar landform characteristics making the modeling of runoff generation challenging(Patel*et al*, 2020).Keeping in view the facts mentioned above, a study was

Comment [HO2]: You said abundant, that means you should have more than one reference

Comment [HO3]: Cite the authors please!

Comment [HO4]: Remove it and keep it in study area section planned to estimate design runoff for various soil and water conservation structures based on annual one-day rainfall using probability distribution fitting as well as artificial neural network (ANN) and gaussian processes regression (GPR).

2. MATERIALS AND METHODS

2.1. Study Area and Rainfall Data

The present study was conducted for the Saurashtra region of Gujarat India), a semi-arid region locatedin western India between 20°30’ to 23° N latitude and 69° to 72° E longitude (Fig.1).

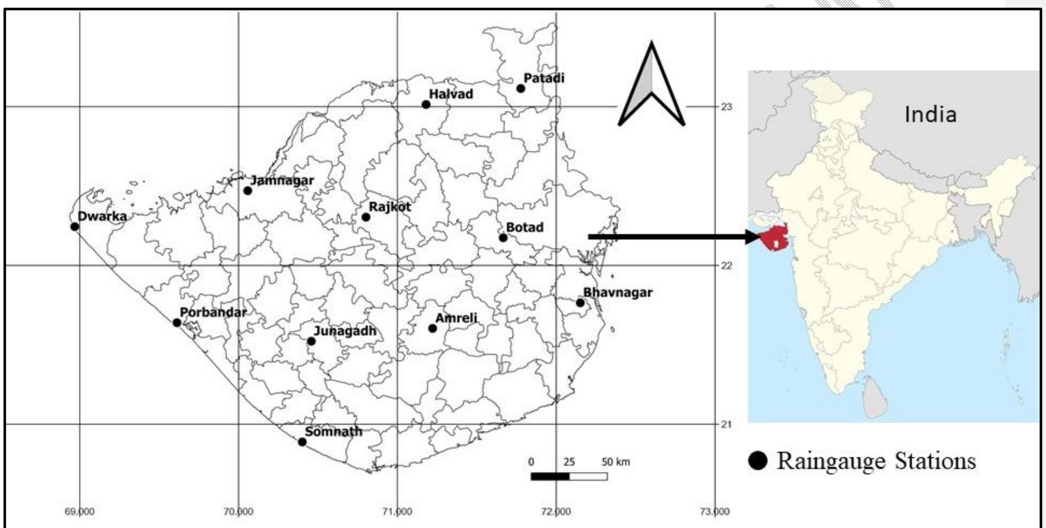

Fig. 1. Study area and location of rain gauge stations

The average annual rainfall over different parts of the region varies widely from 400 to 900 mm. The duration of the rainy season lies between the middle of June to September. The rainfall distribution is uneven and irregular as the region is situated southwest monsoon periphery (Pandya *et al.* 2020). Saurashtra faces high dependence on groundwater along with the absence of major perennial rivers/streamsand recurrent drought-like conditions (Saha *et al.*, 2017; Patel *et al.*, 2020). The daily rainfall data of one station in each of the 11 districts of Saurashtra as shown in Fig. 1 for the year 1981 to 2020 (40 years) were used to obtain a series of data sets of annual one-day maximum rainfall (ADMR) for estimating runoff.

2.2. Probability Distribution Fitting ofADMR

The observed return period of ADMR was calculated by Weibull's plotting position formula (Chow, 1964) and used by Dhupal and Swain (2020).Various nine probability

Formatted: Font: (Default) Times New Roman, 14 pt, Bold

Formatted: List Paragraph, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: (Default) Times New Roman, 12 pt, Bold

Formatted: List Paragraph, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55"

Comment [HO5]: I cannot see these data. Maybe other figure.

Formatted: List Paragraph, Right: -0.54", Space Before: 0 pt, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Tab stops: 1", Left distributions *i.e.* Gamma, Gamma 3 Parameter, Generalized Extreme Value, Log-Logistic, Log-Logistic 3 Parameter, Log-Pearson 3 type, Lognormal, Lognormal (3P) and Normaldistribution were fitted to ADMR.The description of various probability distribution functions is available explained by Sharma and Singh (2010). The distributionswere subjected to three goodness of fit tests*i.e.*Chi-square test, Kolmogorov-Smirnov (K-S) test and Anderson-Darling (A-D) test to find the best fit distribution for each of 11 stations which are discussed in Mandal and Choudhury, 2014 andSingh *et al.* 2018. The best-fit distribution was decided based on the method of ranking distributions by assigning a score from 1 to 9 with distribution having the lowest test statistic value as score 9, second-lowest as 8 and so on. Scores for all three tests were summed up and the distribution with the highest combined score was designated as best-fit(Olofintoye et *al*., 2009).

2.3. Artificial Neural Network (ANN)

Among various ANN models, multi-layer perceptron neural network model (MLP) trained with feed-forward backpropagation algorithm widely used for hydro-meteorological study including rainfall. The architecture of MLP is characterized by the activation function used in each input, hidden and output layer as illustrated in Fig. 2.

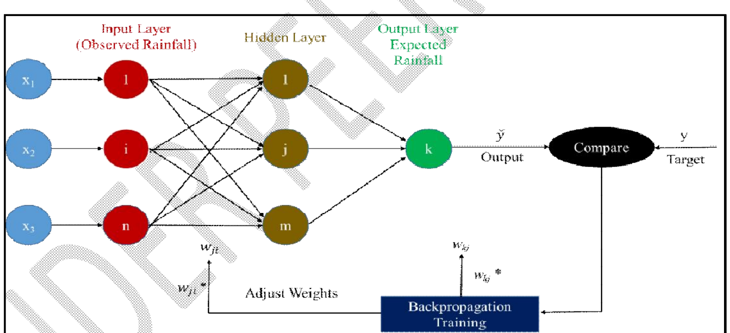

Fig. 2.MLP feed-forward ANN with backpropagation

With $X_i$is input *i.e.* observed rainfall at various return periods, 1 ton are neurons of an input layer, 1 to m are neurons of the hidden layer, k is output neuron and Y is output *i.e.* expected rainfall, w_{ji} and w_{kj} are weights connected hidden & input layers and output & hidden layer.The parameters used during the training of the network are learning rate, Momentum, Number of Neurons, error and Number of Epochs.

2.4. Gaussian Processes Regression (GPR)

The ~~gaussian~~Gaussian process is associated with Gaussian probability distribution, whichdescribes random variables. The Gaussianprocess of f (x) is defined by covariance,

Formatted: List Paragraph, Right: -0.54", Space Before: 0 pt, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: 1", Left

Formatted: Font color: Auto

Formatted: List Paragraph, Right: -0.54", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: 1", Left mean m (x) functions that are a matrix and a vector, respectively, and expresses the distribution between functions.The covariance function leads to the creation of functions with differentdegrees or the different types of continuous structures andprovides the possibility to choose the rightselection. The method Matern covariance function can be used in a majority of which requires specification of onlythe covariance structure (Chang and Chen, 2018). The covariance matrix is required to be a semi-positive definite matrix, and the kernelfunctions fulfill these requirements so they are used to obtain the covariance matrix. The present study incorporated the target response variable as ADMR at various return periods using observed ADMR.

The set of random variables is created which is evaluated at x by GP f (x) where GP is a distribution over functions, *i.e.,*

$$f(x) \sim GP\,(m(x), k\,(x, x^T))$$

Here,m(x) = E[f(x)] is mean function and $k\,(x, x^T) = E[(f(x) - m(x))(f(x^T) - m(x^T))]$ is covariance kernel. The $x = (x_1 \ldots\ldots, x_n)^T$ is the function's index, and $f(x) = (f_1 \ldots\ldots, f_n)^T$ is the function's output, i.e., (x_i, f_i) is a point in R^2. The present study considered common GP, and kernel whichis defined with mean function 0 and squared exponential(SE) covariance function

$$k\,(x_i, x_j) = \sigma^2 \exp\left\{-\frac{1}{2l^2}(x_i - x_j\right\}$$

Where σ^2 and l are hyper-parameters that control the shape of the process; especially, σ^2 controls the amount of variation in f (x) and l, the length-scale parameter, controls the correlation. Several options are available for kernel, however, SE kernel is widely used in GPR (Rasmussen, 2005). The height and amplitude of GP with SE covariance kernel is controlled by $\sigma^2$and l controls the correlation between observations. The covariance matrix K in a multivariate normal distribution is constructed by the kernel as below.

$$K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

Thefinite number n ofrealizations $x = (x_1 \ldots\ldots, x_n)^T$, and corresponding $y = (y_1 \ldots\ldots, y_n)^T$, where $y \triangleq f(x)$ is observed. Thevector x is commonly called the input and represents the location of the process, *i.e.,*observation $y_i = f(x_i)$. The vector y is referred to as the output and is the functionevaluated at location x. The generalization to a multivariate normal distribution is done with f (x) ~ N (0, K). This is possible because marginalizing a

Gaussian distribution istrivial: the resulting distribution is Gaussian and we can ignore the $(x \text{ and } y)$pairs that areunobserved or missing(Forrest and McNicholas, 2020). The hyper-parameters are often estimated to maximize the likelihood ofthe GP.The detailed expressions and calculations are available in Rasmussen and Williams (2006).

The ANN and GPR were perfumed using Weka (Waikato Environment for Knowledge Analysis) machine learning tool developed at the University of Waikato, New Zealand. The detailed calculation steps can be available by Frank*et al.*2016.The performance of developed ANN and GPR in terms of rainfall prediction was tested using the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

2.5. Return Period for Various Soil and Water Conservation Structures

Formatted: List Paragraph, Right: -0.54", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: 1", Left

The return periods of different soil and water conservation structures based on their respective annual one-day maximum rainfall data are 5 years for field bunding, 10 years for terrace outlets and vegetative outlets, 15 years for field diversion, 10 to 15 years for small permanent masonry gully control structures, 25 years for check dams, drainage-line treatment structures, stock water dams, 25 to 50 years for earthen storage dams with natural spillways and 50 to 100 for storage and diversion dams having spillways (Sharma *et al.*, 2015 andDhupal and Swain, 2020).The ADMR for these return periods was evaluated and tabulated.

2.6. RunoffEstimation

Formatted: List Paragraph, Right: -0.54", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Tab stops: 1", Left

Various studies have been conducted to estimate runoff by the SCS Curve number method at various locations, river basins and watersheds in the Saurashtra with curve numbers varying from 68 to 85 for different land use characteristics (Patel *et al.*, 2012;Sonagara *et al.* 2015; Parmar *et al.*,2016andGandhi; Patel 2019). Based on these facts, the design runoff was estimated for various soil and water conservation structures using two curve number values i.e. 70 and 85 for demonstrating the idea.

3. RESULTS AND DISCUSSION~~S~~

3.1. ADMR Based on Probability Distribution Fitting

Formatted: Font: 12 pt, Font color: Auto, Border: : (No border)

Formatted: List Paragraph, Justified, Right: -0.54", Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25", Tab stops: 1", Left

Formatted: Font:

The average annual rainfall based on rainfall data from 1981 to 2020 for various stations under the study ranges from a minimum of 467 mm for Morbi and a maximum of 985 mm for Junagadh. The descriptive statisticsof ADMR for 11 stations of Saurashtra are given in Table 1. It can be observed that themaximum average ADMRwas 167 mm for Junagadh followed by 162 mm for Gir Somnath. While minimum average ADMRwas observed as 96 mm for Botad. In most the cases, a higher average extreme rainfall event corresponds to higher variability and hence higher CV%, which is evident from Table 1. Gebremedhin *et al.* (2017) also confirmed this factduring a studyon probable maximum precipitation (PMP) for Ethiopia.

Table1. Statistical analysis of ~~Annual One Day Maximum Rainfall (~~ADMR~~)~~

Station	AM (mm)	SD (mm)	CV (%)	C_S	C_K
Amreli	110	70	64	1.40	1.25
Bhavnagar	103	56	54	1.06	1.56
Botad	96	45	46	0.97	0.51
Dwarka	112	77	69	1.30	1.75
Gir Somnath	162	107	66	2.31	7.51
Jamnagar	146	95	65	1.14	1.26
Junagadh	167	152	91	4.87	27.68
Morbi	98	47	48	0.79	-0.16
Porbandar	145	122	84	2.99	12.19
Rajkot	120	61	51	1.34	2.28
Surendranagar	107	48	45	0.81	0.13

AM= Arithmetic mean, SD= Standard deviation, CV=Coefficient of variation,
C_S = Coefficient of skewness, C_K = Coefficient of kurtosis

The observed ADMR for return period 1.025 to 41 based on 40 years of data was subjected to variousnine probability distributions. Table 2 depicts the best-fit probability distribution for various stations based on the total score of three goodness of fit tests *i.e.*Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D) and Chi-square test.Out of 11 stations, log-logistic (3P) was found to be the best fit to predict ADMR for the majority of stations (Amreli, Bhavnagar, Botad, Junagadh, Porbandar and Rajkot).During the study in the semi-arid zone of Northwest India, Kumar *et al.* (2018) revealed that each distribution has its pros & cons and one cannot fit in all locations, so examination of best-fit models owing to its peculiar location is essential. Similar here also generalized extreme value distribution was found best-fit two stations *i.e.* Gir Somnath and Jamnagar, lognormal (3P),gamma (3P) and lognormal for one station each.The distributions found to be reliable to predict ADMR were also shown capability to estimate ADMR across India (Babu *et al.*, 2006;Sharma and Singh, 2010; Mandal and Choudhury 2014; Kumar *et al.*, 2018; Dhupal and Swain 2020).

Table 2. Best fit distribution for ADMR

Comment [HO6]: You said that the best-fit probability distribution is there but I cannot see them.

District	Best-Fit Distribution
Amreli Bhavnagar Botad Dwarka Junagadh Porbandar Rajkot	Log-Logistic (3P)
Gir-Somnath	Gen. Ext. Value
Jamnagar	Gen. Ext. Value
Morbi	Gamma (3P)
Surendranagar	Lognormal

Comment [HO7]: You should add the total score of three goodness fit-test and make your table usable

The best fit probability distribution for Junagadh and Rajkot obtained a full score of 27, which reveals that all three tests showed the same probability distribution as the best fit. Except for Morbi, all the remaining 10 stations observed 24 or more score for best-fit probability distributions out of a maximum possible 27, which suggests that obtained best-fit probability distributions were found appropriate by all three goodness of fit tests to predict ADMR. Table 2 also depicts the parameters of best-fit distribution for each station, which represent essential properties of location, scale and shape of a distribution required to predict the expected values. Similarly, in the present study also, a single distribution could not emerge as the best fit for all locations.

3.2. ADMR Based on ANN and GPR

Formatted: Font:

Formatted: List Paragraph, Justified, Right: -0.54", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Tab stops: 1", Left

The performance of ANN and GPRwas evaluated in the testing phase to predict the ADMR for various return periods based on correlation coefficient, mean absolute error and root mean squared error (Table 3).

Table~~.~~ 3. Performance of ~~Gaussian Regression Process (~~GPR~~)~~ and ANN

Station	Correlation Coefficient		Mean Absolute Error, mm		Root Mean Square Error, mm	
	GPR	ANN	GPR	ANN	GPR	ANN
Amreli	0.99	0.52	192	202	213	212
Bhavnagar	0.99	0.52	161	183	174	191
Botad	0.99	0.47	134	135	141	141
Dwarka	0.97	0.55	15	229	21	241
Gir Somnath	0.97	0.50	322	358	369	389
Jamnagar	0.99	0.97	35	293	40	304

Junagadh	0.99	0.45	411	466	494	670
Morbi	0.99	0.86	136	138	143	144
Porbandar	0.99	0.52	337	371	401	420
Rajkot	0.99	0.78	190	203	206	212
Surendranagar	0.99	0.47	148	141	155	148

The higher correlation coefficient between observed and predicted rainfall ranging from 0.97 to 0.99 for GPRwas observed as compared to ANN (0.45 to 0.86). The mean absolute error in rainfall prediction was lower i.e. from 15 to 411 mm for GPR i.e. as compared to ANN (135 to 466 mm) for various stations. Among all stations, only Surendranagar was observed with lower MAE in ANN (142 mm) as compared to GPR (148 mm). While RMSE value of GPR and ANN for stations Amreli, Botad, Morbi, Rajkot and Surendranagar was very close to each other with the difference being less than 10 mm. The difference in RMSE between GPR and ANN was low i.e. 16 to 20 mm for Bhavnagar, Gir Somnath and Porbandar while for Dwarka, Junagadh and Jamnagar it was as high as 175 to 219 mm.These findings are also supported by a study by Sahraei *et al.* (2021) disclosing that ANN underestimated the events with high maximum values and overestimated the events with low maximum values.Hence, considering all three performance indicators i.e. correlation coefficient, mean absolute error and root mean square error, GPRperformed better than ANN to predict the ADMR for various stations. Sudheer *et al.* (2003) during the study of river flow simulation reported that ANN models suffer from the weakness in predicting extreme events unless they are trained by similar extreme events. As in the present case also, the observed values of rainfall obtained using Weibull's plotting position was ranging from 1.025 years to a maximum of 41 years, but expected rainfall for 50 and 100 years return period was also evaluated. Another reason to point out is characteristics "sigmoidal" function of ANN are to be bounded and increase monotonically to the variability induced by extreme values. On the contrary, the parsimonious structure of GPR does not allow overfitting in such cases(Chang and Chen, 2018).The ANN might perform better in this case if data of more than 40 years are used for the analysis. Mishra and Kushwaha (2019) also observed the satisfactory performance of the Gaussian process regression model in precipitation forecasting.Hence, considering the superiority of GPR over ANN the runoff was also estimated by ADMR obtained from GPR along with best fir distribution.

3.3. Runoff Estimation Using ADMR

Comment [HO8]: Fit??

Formatted: List Paragraph, Right: -0.54", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.55", Tab stops: 1", Left

The expected ADMR for various return periods based on best-fit probability distribution and GPR is given in Table 4.

Table~~.~~ 4. Design rainfall for various return periods (mm)

Station	Method	5	10	15	25	50	100
Amreli	~~PBF~~PDF	143	201	247	315	443	625
	GPR	126	147	167	207	308	510
Bhavnagar	PDF	140	175	196	226	271	324
	GPR	116	132	147	179	259	417
Botad	PDF	130	154	173	198	238	286
	GPR	106	118	130	155	216	338
Dwarka	PDF	163	211	241	277	328	381
	GPR	130	153	175	220	332	557
Gir Somnath	PDF	214	281	326	386	481	592
	GPR	189	224	258	328	501	847
Jamnagar	PDF	209	268	306	353	419	490
	GPR	167	194	221	274	409	677
Junagadh	PDF	206	255	288	331	398	477
	GPR	207	257	307	407	656	1156
Morbi	PDF	133	163	180	201	227	253
	GPR	108	120	132	157	217	339
Porbandar	PDF	194	264	313	381	498	649
	GPR	177	218	258	339	541	946
Rajkot	PDF	157	197	223	258	315	383
	GPR	134	152	170	206	297	477
Surendranagar	PDF	142	173	192	215	247	279
	GPR	117	130	142	168	231	358

PDF:- Probability distribution fitting

The design rainfall based on GPR for 5, 10, 15 and 25 years return period was found less as compared to that of the best-fitted probability distribution for all the stations except Junagadh. For 50 years return period, Dwarka, Gir-Somnath, Junagadh, Morbi and Porbandar observed a higher value of expected rainfall for GPR as compared to that of best-fit distribution fittings. While for 100 years return period, the rainfall by GPR was found higher than by best-fit probability distribution for all districts except Amreli. Junagadh was the only station for which rainfall predicted by GPR was higher than by probability distribution for all the six return periods considered. The observed average ADMR based on 40 years of data was highest for Junagadh as compared to all the districts (167 mm), in the case of 100 years return periods, expected rainfall was naturally on the higher side. For such higher rainfall

Comment [HO9]: For Morbi : expected rainfall for GPR<expected rainfall for PDF

values, the difference between the expected value by a probability distribution and GPR was also high as compared to that of lower return periods. Thus, it can be interpreted that for higher observed ADMR events and higher return periods, GPR estimates are higher as compared to probability distribution and vice-versa. The estimated ADMR for various return periods serves as input for the hydrologic design of different soils and water conservation structures. Babu *et al.*2006 demonstrated the capability of ADMR for the efficient design of a masonry check dam in Bankura district of West Bengal based onprobability distribution fitting.

The design runoff was estimated for various soils and water conservation structures using two curve number values i.e. 70 and 85 (Table 5 and Table 6). However, runoff can be estimated using any curve number by using ADMR in Table 4. The expected runoff for filed bunding based on best-fit distribution was ranging from a minimum 54 mm (Botad) to 123 mm (Gir Somnath) for CN70 and from 88 mm (Botad) to 168 mm (Gir Somnath) for CN85. Considering GPR, the expected runoff was ranging from 46 mm (Botad) to 131 mm (Junagadh) for CN70 and from 66 mm (Botad) to 161 mm (Junagadh) for CN85. Similarly for Storage and Diversion Dams having spillways considering 100 years return period, the expected runoff based on best-fit distribution was ranging from a minimum 157 mm (Morbi) to 534 mm (Porbandar) for CN70 and from 206 mm (Morbi) to 590 mm (Porbandar) for CN85. Considering GPR, the expected runoff was ranging from 254 mm (Botad) to 1060 mm (Junagadh) for CN70 and from 290 mm (Botad) to 1103 mm (Junagadh) for CN85 for Storage and Diversion Dams having spillways. The design runoff for structures like Terrace outlets and vegetative outlets, Field diversion, Small permanent masonry gully control structures, Check Dams, Drainage-line treatment Structures (DTS), Stock Water Dams and Earthen Storage Dams with Natural Spillwayscan also be observed in Table 5 and 6.

Comment [HO10]: I cannot see them in the table

Table 5. Estimated runoff for 5, 10 and 15 years return period (mm)

Station	RP 5 Years				RP 10 Years				RP 15 Years			
	CN70		CN85		CN70		CN85		CN70		CN85	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Amreli	64	62	100	85	112	78	156	104	151	96	200	123
Bhavnagar	61	53	97	75	89	66	130	90	107	79	151	105
Botad	54	46	88	66	72	55	110	77	88	65	129	89
Dwarka	80	65	119	89	120	84	165	110	147	103	194	131

Gir Somnath	123	115	168	144	183	147	234	178	224	179	278	211
Jamnagar	118	96	163	123	171	120	221	149	206	144	258	175
Junagadh	115	131	160	161	159	177	208	210	189	224	240	259
Morbi	56	47	91	68	80	57	119	79	94	66	136	90
Porbandar	106	105	149	133	167	141	216	172	212	178	265	211
Rajkot	75	68	114	92	108	83	152	109	131	99	177	126
Surendranagar	63	54	100	76	88	64	129	88	104	75	147	100

(1) based on best-fit distribution, (2) based on GPR

Table 6. Estimated runoff for 25, 50 and 100 years returns period (mm)

	RP 25 Years				**RP 50 Years**				**RP 100 Years**			
Station	**70**		**85**		**70**		**85**		**70**		**85**	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Amreli	214	132	267	162	335	225	393	260	511	421	574	460
Bhavnagar	133	107	180	135	174	179	224	212	222	330	276	368
Botad	109	85	153	111	144	139	192	170	187	254	238	289
Dwarka	179	143	230	174	226	249	280	284	276	467	332	506
Gir Somnath	280	244	337	279	371	412	431	451	479	753	541	795
Jamnagar	249	194	304	227	312	322	370	359	380	585	440	626
Junagadh	229	320	283	357	292	564	349	605	367	1060	427	1103
Morbi	111	87	155	113	134	141	181	172	157	255	206	291
Porbandar	276	255	332	291	388	451	448	491	535	851	598	894
Rajkot	162	131	211	161	214	215	267	249	278	389	334	427
Surendranagar	123	96	169	124	151	154	200	185	181	274	232	310

(1) based on best-fit distribution, (2) based on GPR

From the economical design point of view, among expected runoff by probability distribution function and GPR, the lower of the two values should be considered for the structure design. However, in the situation where finance is not much constraint, the higher of the two can be used. This study has demonstrated the potential of machine learning technic like the Gaussian regression process(GPR) in addition to probability distribution fitting to standardize the hydrologic design of various structures for a particular region using ADMR. The study can serve as a ready reckoner for design engineers, planners and policymakers for efficient and economic structure design as well as working out cost economics for various soils and water conservation interventions during project proposal preparation. The replication of such study in a region would help in planning, designing and management of different types of hydraulic structures for effective utilization of available water resources.

4. CONCLUSION~~S~~

Formatted: List Paragraph, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: (Default) Times New Roman, 14 pt, Font color: Black

The probability distribution fitting of ADMR revealed that out of 11 stations, log-logistic (3P) was found best-fit for 6 stations, generalized extreme value distribution for two stations and lognormal ~~(3P)~~, and gamma (3P) ~~and lognormal~~ for one station each. The GPR has shown superior performance to predict ADMR at various return periods as compared to ANN based on correlation coefficient, mean absolute error and root mean squared error. The expected ADMR based on GPR and best-fit distribution was evaluated for various structures incorporating return periods 5 to 100 years. The design runoff estimation for CN 70 and CN85 for various soils and water conservation structures using GPR and probability distributions was demonstrated. The results from this study will be useful to formulate the efficient and economic design of various soils and water conservation structures and help in utilizing the available rain and land resource to their potential.

Comment [HO11]: 7 according to Table 2. (Amreli Bhavnagar Botad Dwarka Junagadh Porbandar Rajkot)

REFERENCES

Comment [HO12]: Use the author guidelines to correct the reference style, please

Babu, R., Mishra, P., Mazumdar, A. and Roy, D. 2006. Probability analysis of rainfall of Bankura for design of soil and water conservation structures. *Journal of agricultural engineering*., 43(1):22-29.

Bitterman, P., Tate, E., Van Meter, K. and Basu, N. 2016. Water security and rainwater harvesting: A conceptual framework and candidate indicators.*Applied Geography*., 76(6):75-84.

Chang, W. and Chen, Xi. 2018. Monthly rainfall-runoff modeling at watershed scale: a comparative study of data-driven and theory-driven approaches. *Water*, 10:1116.

Chow, V.T. 1951. A general formula for hydrologic frequency analysis. *Transactions American Geophysical Union*., 32(2): 231-237.

Darji, M., Dabhi, V. and Prajapati, H. 2015. Rainfall forecasting using neural network: A survey. *In:* Proceedings of an International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India. p. 706-713.

Dhupal, G. and Swain, S. 2020. Estimation of probable maximum one day duration rainfall for Khurda region. *International Journal of Current Microbiology and Applied Sciences*., 9(5): 2450-2462.

Forrest, P. and McNicholas, P. 2020. Detecting British Columbia Coastal Rainfall Patterns by Clustering Gaussian Processes. *Environmetrics*. 31 (8).

Frank, E., Hall, M. A. and Witten, I. H. 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Gandhi, F., R. and Patel, J. N. 2019.Estimation of surface runoff for sub-watershed of Rajkot district, Gujarat, India using SCS – curve number with integrated geo-spatial technique.I*nternational Journal of Engineering and Advanced Technology*, 8, 33-41.

Gebremedhin, Y. G., Quraishi, S. and Itefa, H. 2017. Development of one day probable maximum precipitation (PMP) and isohyetal map for Tigray Region, Ethiopia. *International Journal of Human Resources Development and Management.*, 32:101-122.

Ghasemi, P., Karbasi, M., Zamani, N., Alireza, S., Tabrizi, M. and Azamathulla, H. 2021. Application of gaussian process regression to forecast multi-step ahead SPEI drought index. *Alexandria Engineering Journal.*, 60:5375–5392.

Gundalia M. J. 2016. Modelling Runoff Using Modified SCS-CN Method for Middle South Saurashtra Region (Gujarat-India). (Doctoral Dissertation, Gujarat Technological University Ahmedabad).

Kumar, M., Anurag, Kumar, S., Singh, B., Singh, R. 2018. Estimation of return period probability of maximum rainfall in a day for Hisar, Haryana, India. *JournalAgrometeorology.*, 20. 128-134.

Mandal, S. and Choudhury, B 2014. Estimation and prediction of maximum daily rainfall at Sagar Island using best fit probability models. *Theoretical and applied climatology*, 121:87-97.

Mishra, N. and Kushwaha, A. 2019. Rainfall prediction using gaussian process regression classifier. *International Journal of advanced research in computer science engineering and information technology.*, 8(8): 392-397.

Nahvi, A., Daghighi, A., and Nazif, S. 2018. The environmental impact assessment of drainage systems: a case study of the Karun river sugarcane development project. *Archives of Agronomy and Soil Science.*, *64*(2):185-195.

Olofintoye, O., Sule, B.F., Salami, AW. 2009. Best–fit probability distribution model for peak daily precipitation of selected cities in Nigeria. *Nigeria Science Journal.*, 2(3):1–12.

Otim, D., Smithers, J., Senzanje, A. and Antwerpen, R. 2019. Design norms for soil and water conservation structures in the sugar industry of South Africa. *Water S.A.*, 45(1):29-40.

Pandya P. A., Kumarkhaniya, R., Parmar, R. and Ajani, P. 2020. Meteorological drought analysis using standardized precipitation index. *Current World Environment.*, 15(3):477-486.

Paola De, F., Ranucci, A. and Feo, A. 2013. Antecedent moisture condition (SCS) frequency assessment: A case study in southern Italy.*Irrigation and Drainage*., 62, 61–71.

Parmar, H. V., Mashru, H. H., Vekariya, P. B., Rank, H. D., Kelaiya, J. H., Pardava, D. M., Patel, R. J. and Vadar, H. R. 2016. Establishment of rainfall - runoff relationship for the estimation runoff in semi-arid catchment. *AGRES – An International e-Journal.*, 5(1):60-67.

Patel, G. R., Patel, R. J. and Chalodia, A. L. 2012. Agricultural planning and sustainable development of ungauged watershed area using remote sensing and GIS. *International Journal of Management Science and Engineering Management.*, 4(1):23-28.

Patel, P., Saha, D., and Shah, T. 2020. Sustainability of groundwater through community-driven distributed recharge: an analysis of arguments for water scarce regions of semi-arid India. *Journal of Hydrology: Regional Studies*., (29).

Rasmussen, C.E. and Williams, C.K.I. 2006. Gaussian Processes for Machine Learning; The MIT Press: Cambridge, MA, USA, 2006; ISBN 026218253X.

Saha, D., Marwaha, S., Mukherjee, A. 2017. Groundwater resources and sustainable management issues in India. *In:* Saha, D., Marawaha, S., Mukherjee, A. (Eds.), Clean and Sustanable Groundwater in India. Springer, Singapore, p.4551–4556.

Sahraei, A., Chamorro, A., Kraft, P. and Breuer, L. 2021. Application of machine learning models to predict maximum event water fractions in streamflow. *Front. Water*, 3:1-21.

Sharma, K., Singh, A. and Dubey, S. 2015. Analysis of one day probable maximum precipitation for designing soil and water conservation structures in Agra, U.P. *Jornal ofAgrometeorology*., 17(2):268-270.

Sharma, M. and Singh, J.B. 2010. Use of probability distribution in rainfall analysis. *New York Science Journal*., 3(9):40-49.

Singh, A., Singh, V. and Byrd, A. R. 2018. Computation of probable maximum precipitation and its uncertainty. *International Journal of Hydrogen Energy*, 2(4):504-514.

Singh, G., Dinesh, D., Kakade V. D., Bhatnagar, P.R. and Pande, V.C. (2019). Precipitation probability and water budgeting for crop planning in central Gujarat. *Journal of Agrometeorology*., 21(3):392-396.

Singh, R., Singh, K. and Bhandarkar, D.M. 2017. Hydrologic design parameters database for water harvesting structures in Madhya Pradesh.(EDs. V. P. Singh, S. Yadav and R. N. Yadava). pp. 161-74, Hydrologic Modeling, Water Science and Technology Library, 81.

Songara, J., Kadivar, H.T., Joshipura, N. and Prakash, I. 2015. Estimation of surface runoff of Machhu dam III catchment area, Morbi, Gujarat, India, using curve number method and GIS.*International journal of scientific research*., 3:2038-2043.

Sudheer, K. P., Nayak, P. C. and Ramasastri, K. S. 2003. Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, 17:677–680.

USDA-SCS (1986). "Urban hydrology for small watersheds." U. S. Department of Agriculture, Technical Release No. 55.