Biological Forum – An International Journal 16(7): 341-347(2024)

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Probability Distributions Models to Optimize the Rainfall in Karnataka State

Vinushree R., Ashish Baluni*, Megha J. and Chetan Mahadev Rudrapur
Department of Agricultural Statistics,
University of Agricultural Sciences, Dharwad (Karnataka), India.

(Corresponding author: Ashish Baluni)*
(Received: 01 May 2024; Revised: 27 May 2024; Accepted: 19 June 2024; Published: 15 July 2024)
(Published by Research Trend)

ABSTRACT: The daily rainfall data (1988-2018) were analysed to extrapolate maximum rainfall in Dharwad district of Karnataka state, India. Newer run-off model was formulated to study the different cropping patterns@ different significant levels. Absolute rainfall data were divided into 23 sets *viz.* one-annual and four seasons (June to September) were Standardized based on different meteorological weeks (23-39rd SMW). The model findings showed that, the annual maximum daily rainfall IQR ranged between (23.14-83.44) mm indicated with large amount of fluctuation of rainfall because due to periodic changing of the weather parameters and global warming. Asper the models diagnostic checking, the runoff- probability forecasted distribution model was found to be best fit to optimize the absolute daily rainfall. Our formulated model will be very useful for the agricultural scientists for designing of cropping patterns and also it can be served as navigation tool for the meteorologists, agricultural policy planners and researchers for the holistic development of soil conservation strategies and irrigation purpose in the developing countries.

Keywords: Maximum daily rainfall, Probability distributions, Kolmogorov-Smirnov test.

INTRODUCTION

Rainfall is one of the most important natural resources for the growth of agricultural crops. The maximum rainfall is related to natural systems and unpredictable seasons. Paucity of mathematical/statistical computer-aided predictive modelling techniques makes the forecasting of rainfalls (uncertainties and likelihoods), the major biggest problems facing the world today. Rainfall and temperature variations are caused by environmental factors and are affected by the cyclone's periodic effects. One of the difficult worldwide challenges of the day is how to accurately identify rainfall and other meteorological conditions. The effectiveness of agricultural cultivation and irrigation will be determined by on time frame. The long-term study of rainfalls is necessary to carry out planning and cultivating crops. However, success and failure depends on various rainfall conditions. It is closely correlated with the weather forecast. Imbalanced weather conditions had a significant impact on human resources, farming, management and crop production. The weather is extremely complicated, making it difficult to choose the best choice for crop production. Due to the negative impacts of climatic conditions, the SARS-CoV-2 pandemic and other bacterial infections have spread throughout the world, discouraging farmers and agricultural specialists from focusing on managing agricultural resources, such as imports and exports to a global level (distribution of seeds, fertilisers and agricultural equipment's etc. Accordingly, the explanation of heavy rainfall encourages taking into account unexpected events that make projecting future output and economic losses. For the management of crops and the adoption of modern agricultural policies for changing the cropping pattern to a greater extent, simulation data driven prediction support modelling is absolutely critical. The present analytical study will help the prediction of rainfall and management of water and natural resources. This navigated formulated model will be the greatest advantage for forecasting weather parameters, floods, cyclonic seasonal variations and drought prone areas at national level. This study attempt to determine to formulate various probability distribution models to optimize the maximum daily rainfall.

MATERIALS AND METHODS

Thrust area. Dharwad district was purposively selected for the study, it is located in the north-western part of Karnataka, co-ordinated @ 15°27′30″ North and East @ 75°00′30″. As experienced tropical wet climate, (the similar ≈741 meter above the sea level). The boundary was encompassed with a geographical area of 4263 sq. kms. South-west monsoon is the most common which can be received maximum rain fall. An average annual rainfall was ≈ 864 mm (over five years). Very less rainfall was recorded in winter instead of summer season. Humid/dry weather in Feb to early June month, the highest quantity during precipitations with mean temperature (24.1°C). The hot climate in the summer (April–May) and nice throughout the year. The hot season begins from March to Jun with the highest temperature of (38ºC) and lowest temperature (14ºC) during the month of December (freezing month). After SARS-CoV-2 pandemic attacked, all weather parameters have been changed, periodically hitting cyclones and got maximum rain fall in the year 2021 -2022 (regular rainfall > 1800mm). Under these conditions, the net crop sown area is very less in Dharwad district in 2020-2022; it will have a significant impact on both economic security and crop yields. The accurate likelihood estimation of rainfall is very essential for estimation of cause and effect relationship between the climatic conditions and cropping patterns at national level. In this practical manner, we have formulated newer run off -model for absolute estimation and its likelihoods. A preliminary stage, the sample data were used to demonstrate our projected models. The selection of sample has drawn from the total geographical areas it was determined by the following eqn (1)

$$D = max_{1\leq i\leq n}(FX_i) - \frac{i-1}{n}, \frac{i}{n} - FX_i \qquad (1)$$

Where, X_i = Random sample, i= 1, 2,…, n ;

CDF =F(x)= $\frac{1}{n}$[number of observations ≤ x] .

This test was used to decide the samples drawn from hypothesized distribution. Finally, we have formulated probability distribution models to optimize the exact rainfall. The above methodology will also be used for studying the probability distribution pattern for other disciplines. In case of model formulation, we used daily rainfall data besides with different weather parameters between (1988-2022), the data was collected from the Department of Agricultural Meteorology, Government of Karnataka and University of Agricultural Science , Dharwad. The primary and secondary data were used for demonstration of the formulated model.

Model formulation

$$Y_{ijk} = \left(\alpha_i + \beta_{jk}\right)^{t_k} + \epsilon_{ijk} \qquad (2)$$

Where,

Y_{ijk} = Observed rail fall for i^{th} village j^{th} district k^{th} cluster

α_i = Effect of i^{th} village from rain fall

β_{jk} = Interaction effect of rail fall on j^{th} districts k^{th} cluster

t_k = The rain fall seasons at k^{th} clusters

ε_{ijk} = Error associated with i^{th} village j^{th} districts k^{th} cluster

In equation (1) we predicated the random process of and correlation of rainfall in selected villages. The average rainfall was estimated for building the interaction effects with noisy data sets in eqn (2)

$$Y_{ijk} = \left(\frac{1}{1-\alpha_i+\beta_{jk}}\right)^{n_i} + \epsilon_{ijk} \qquad (3)$$

n_i = Number of years attributed to the average rainfall occurs in particular seasons

For the elimination of noisy data below the average rainfall in the particular districts, we build the models in following eqn (4)

$$Y_{ijk} = \left(\frac{1}{1-\alpha_i+\beta_{jk}}\right)^{n_i} + \epsilon_{ijk} \qquad (4)$$

Run- off Predictive model (RPM). The runoff-model is versatile and robust analytical algorithmic predictive model, that was confined to estimates the absolute rain fall in the selected areas. Mainly we substituted infiltrations, evapotranspiration, and depression storage and drainage losses in particular agriculture crops. Evolving of runoff model which is consisted too- many factors of random variables 'Xi'. The random variables 'Xi' is normally distributed (Gaussian distribution) with mean μ, and variance σ^2.

The cumulative estimation of rain fall in the Gaussian distribution attributed with rv's is estimated in the eqn (5)

$$Y_{i=0,1} = \frac{1}{\left(1+Exp^{-\frac{1}{2}(\alpha_i+X_i\beta_{jk})}\right)} \qquad (5)$$

The most widely used non parametric method is the kernel density estimates .We describing the initial losses of random variables {X_1, X_2, …. X_n} with kernel function k(.), the probability density function (PDF) can be estimated in the eqn (6)

$$Y_{ijk} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \qquad (6)$$

where 'h' is smoothing factor, known as the bandwidth and K(.) is the Gaussian Kernel density with distribution $X_i \sim N(\mu, \sigma^2)$

$$K(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \qquad (7)$$

The bandwidth is the most important characteristics of a Kernel density estimates (KDE) with a strong association between shape and scale parameters of PDF

$$RF_{Optimal} = 1.06 * Min\left\{\sigma \frac{IQR}{1.34}\right\} * n^{1/2} \qquad (8)$$

Where, the IQR is the Interquartile range of rainfall in the selected years @ time 't' and 'σ' is the sample standard deviation.

In case of rainfall-runoff modelling, the loss parameter is one of the most important parameter, which can describe the total amount of rainfall. Which was estimated by using infiltration, evapotranspiration, interception, depression storage and transmission losses etc. The several parameters were considered for the formulation of modelling eqn (9)

$$Runoffmodel = \left[t_{\alpha/2} * Baserainfall\left(\frac{N}{1+exp^{(\alpha+\beta_0 x_0+\beta_1 x_1+\cdots\beta_n x_n)}}\right)\right] \qquad (9)$$

Where $t_{\alpha/2}$ = hypothetical table value (1.96)

N = Number of selected waste hr/soil parameters

β_0, β_1, β_2 … β_n is the coefficients of random variables of x_i (Soil moisture, water wastage from drainage, ground water layer, soil condition/types of soil, turbidity of the soil etc).

These parametric distributions rely on the assumption that the population (which the observed data sample belongs to) has an underlying parental distribution. These results were observed from the kernel density function. A nonparametric approaches also been used to correlated prior and posterior information's of rainfall. However, we substituted both assumptions that have been demonstrated by underlying probability distribution. This method will be, relies heavily depends on the absolute rainfall data and its empirical grouping of individual bins. This model based study has compared for both parametric and nonparametric distributions @ initial losses of rainfall and outlier effects on periodicity of cyclones.

$$f(x) = (n - k + 1) + a_i \sum_{i=1}^{ni} \beta_{ij} + \epsilon_{ij} \qquad (10)$$

Where,

f (x) = The function f(x) describes the absolute rainfall yield with (x) random process

n = Number of weather parameters responsible for RF

k = Number of success of geographical areas selected for likelihood estimation

a_i = Regression intercept (point of infelcxion)

β_{ij} = Regression coefficient with respect to different parameter under concern

(At i^{th} traits with j^{th} areas)

ε_{ij} = Residual error with i^{th} traits with j^{th} areas.

(α = k + 1) ; (β = n – k + 1)

Example

α = k + 1 = (7 + 1) = 8, β = 10 – 7 + 1 = 4

The random behaviour of proportion of runoff rate was modelled by β-distribution. It provides powerful tool quantify the likelihood estimation tasks. The probability density function is

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \ \alpha > 0 ; \beta > 0 ; 0 \leq x \leq 1 \qquad (11)$$

Beta distribution is the conjugate of Binomial distribution

$$f(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \text{ BD} \qquad (12)$$

$$g(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{x-1} (1-\theta)^{n-x} \qquad (13)$$

RESULTS AND DISCUSSION

One day maximum daily rainfall corresponding date for the period of 31 years (1988-2018) is presented in the Table 1. The maximum (83.44 mm) and minimum (23.14 mm) annual one day maximum rainfall was recorded during the year 2000 (11^{th} June) and 2016 (28^{th} June) respectively. This indicates that the most fluctuations were observed during the decade 2000-11. The average for these 31 years rainfall was found to be 41.23 mm. It was also observed that 14 years (45.16%) received one day maximum daily rainfall above the average, however, no general trend in rainfall occurrence was observed during the study period. Fig. 1 shows the variation in one day maximum daily rainfall. The distribution of one day maximum rainfall received during different months in a year is presented in Fig. 2. From the Fig. 2, it can be seen that June received the highest amount of one day maximum rainfall (23%) followed by July (16%), September (16%) and October (16%) (Singh *et al.*, 2012).

The data was classified into 23 sets as mentioned in the Table 2 *viz.*, 1 annual, 1 seasonal, 4 seasonal months and 17 seasonal SMW's to study the distribution pattern of rainfall at different levels.

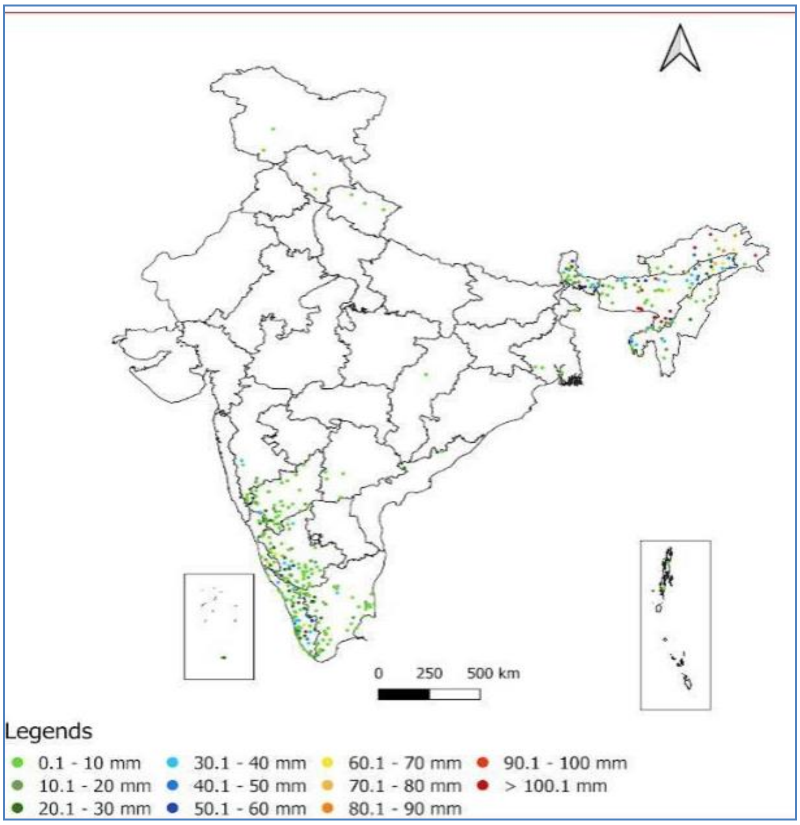

Fig. 1. Geographical distribution of rainfall in Karnataka State.

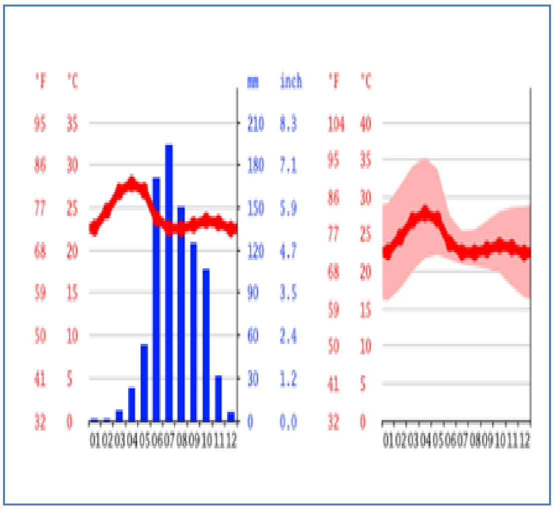

Fig. 2. Absolute mean earth surface temperature in Dharwad District.

Table 1: One day maximum daily rainfall for the period of 1988 to 2018.

Year(t)	Days	Rainfall (mm)	Year(t)	Days	Rainfall (mm)
1988	Jul-18	33.29	2007	Sep-17	50.01
1989	Sep-20	60.66	2008	Mar-22	43.99
1990	May-27	23.64	2009	Oct-01	65.49
1991	Jun-06	62.38	2010	Sept-23	37.25
1992	Nov-18	36.88	2011	Aug-19	42.71
1993	Oct-18	46.59	2012	Nov-01	42.86
1994	Apr-12	50.5	2013	Jun-05	34.61
1995	Jul-07	26.57	2014	Oct-25	40.02
1996	Sep-05	45.55	2015	Aug-17	33.15
1997	Jun-14	41.3	2016	Jun-22	23.14
1998	Aug-21	38.69	2017	Sep-28	29.06
1999	Jul-21	31.43	2018	Jun-02	27.06
2000	Jun-11	84.44	2019	Jul-23	34.22
2001	Oct-07	27.22	2020	Aug-21	40.25
2002	Aug-01	66.9	2021	Oct-15	55.22
2003	Oct-5	23.46	2022	Sep -17	76.52
2004	Jul-12	31.31	-	-	-
2005	Jul-25	41.49	-	-	-
2006	Jun-9	36.47	-	-	-

3.91 10.85 meters below ground level in 2015 to 3.91 meters below ground level in 2019.

Table 2: Parameters of the best fit probability distributions for maximum daily rainfall.

Study Period	Range	Best fit	Parameters		
			Shape (k)	Scale (β)	Location (μ)
Annual	Jan 1- Dec 31	Log-normal	0.322	3.665	-
Seasonal	Jun1- Sep30	Log-normal	0.341	3.595	-
June	Jun 1- Jun 30	Log-normal	0.407	3.27	-
July	Jul 1- Jul 31	GEV	9.294	21.492	0.033
August	Aug 1- Aug 31	Log-normal	0.37	3.041	-
September	Sep 1- Sep 30	Log-normal	0.617	2.971	-
23rd SMW	Jun 4- Jun 10	GEV	7.99	12.637	0.115
24th SMW	Jun 11- Jun 17	Log-normal	0.596	2.788	-
25th SMW	Jun 18- Jun 24	Log-normal	0.739	2.472	-
26th SMW	Jun 25- Jul 1	GEV	5.545	9.447	0.012
27th SMW	Jul 2- Jul 8	Weibull (2P)	1.938	14.914	-
28th SMW	Jul 9- Jul 15	Gamma (2P)	2.374	6.749	-
29th SMW	Jul 16- Jul 22	Log-normal	0.721	2.486	-
30th SMW	Jul 23- Jul 29	Gamma (2P)	1.644	9.838	-
31st SMW	Jul 30- Aug 5	Log-normal	0.796	2.334	-
32nd SMW	Aug 6-Aug 12	Weibull (2P)	1.633	13.896	-
33rd SMW	Aug13-Aug31	Log-normal	0.713	2.29	-
34th SMW	Aug 20- Aug26	Log-normal	0.759	2.327	-
35th SMW	Aug 27-Sep2	Weibull (2P)	1.514	13.81	-
36th SMW	Sep3-Sep9	Gamma (2P)	1.384	6.487	-
37th SMW	Sep10- Sep16	Weibull (2P)	1.05	10.293	-
38th SMW	Sep17- Sep23	Weibull (3P)	0.924	12.148	0.471
39th SMW	Sep24- Sep30	Gamma (2P)	1.884	8.255	-

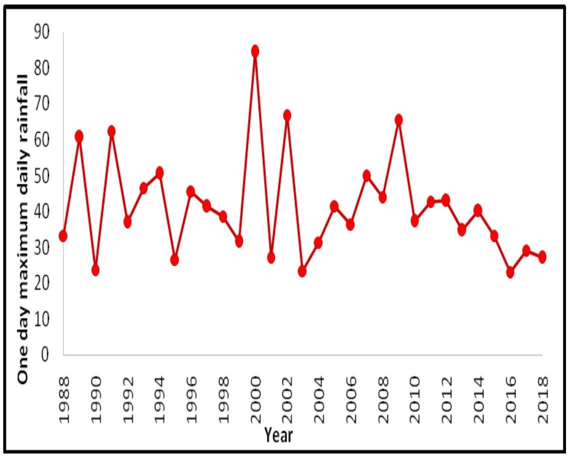

Fig. 3. Year wise annual maximum daily rainfall (in mm).

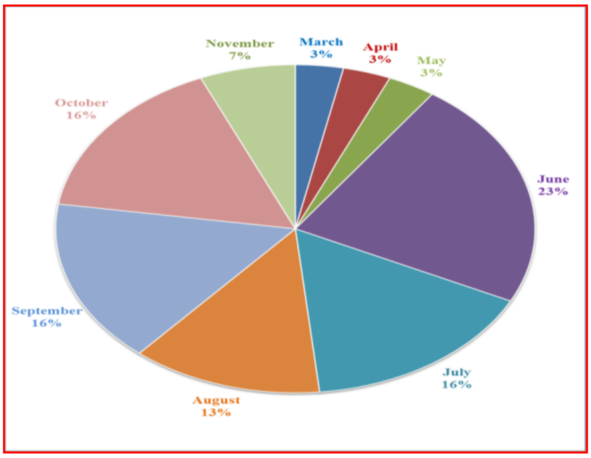

Fig. 4. Distribution of one day maximum annual rainfall in a year.

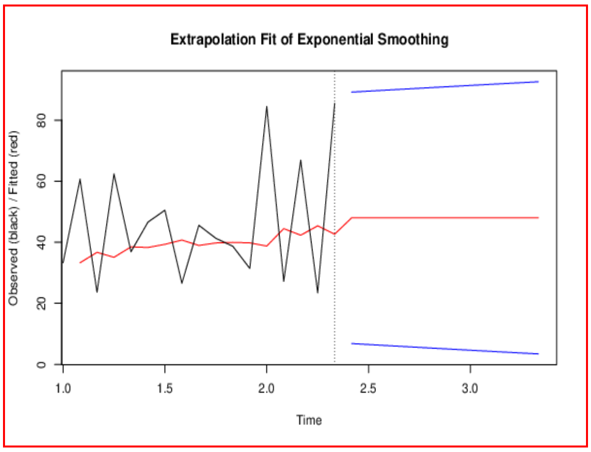

Fig. 5. Extrapolation of minimum rainfall by exponential smoothening by Gaussian.

From the formulated model the analysis observed that majority of the data sets Log-normal distribution was found to be the best fit probability distribution. Similar results were obtained by Manikandan *et al*. (2011). Further, majority of the study periods were observed to be scale dominated which showed large variation in the distribution of rainfall. The best fit probability distributions for maximum daily rainfall on different sets of data were identified based on the criteria of Kolmogorov-Smirnov test. A similar study was conducted by Nassif *et al*. (2021**)**. The distributions such as Normal, Log-normal, Gamma (2P), GEV and Weibull (2P) were found to be the best fit for all the study periods. During the month of August, 25th, 28th, 35th and 38th SMW's, Weibull (3P) distribution fitted well along with the distributions mentioned above. For annual, seasonal, June, August and September study periods, Log-normal was found to be the best fit probability distribution. For July month, GEV distribution fitted well with the lowest test statistic value and the highest p-value. For the 24th, 25th, 29th, 31th, 33th and 34th SMW's, Log-normal was found as the best fit probability distribution, for 28th, 30th, 36th and 39th SMW's, Gamma (2P) was found as the best fit probability distribution, for 27th, 32nd, 35th and 37th SMW's, Weibull (2P) was found as the best fit probability distribution, for 23rd and 26th SMW's, GEV was found as the best fit probability distribution and for 38th SMW, Weibull (3P) was found as the best fit probability distribution with the lowest test statistic value and the highest p-value. In this formulated model we demonstrated real data sets at possible iteration to optimize by means of combining state variables. We start with above formulated predictive models, which are supposed to have been developed in design phase of RF estimation. As an example, eqn (10) parameter of the model we can optimize by means of historical data sets as part of the optimizing techniques, as model construction we substituted different state variables like too many weather parameters, RF geographical area, runoff surface flow, wind blow, earth surface temperature , wind pressure and ocean current etc. Due to outlier state variable model (heavy rainfall due to occurrence of periodic cyclone, flood, drastic changes of climate for heavy RF, biotic intervention etc), the model was rebooted and lead with large inherent uncertainty (estimated errors has deeply penetrated to reproduce imbalance forecasted figures).The above critical circumstances, the model can able to allow the substitution parameters to render the absolute forecasted figures (set of forecasted quality) for usual interpretations. For a regression model eqn () without unobserved components, such as predicted model, the error $\epsilon M_k(i+m)$statistics are the same as forecast error $\epsilon F_k(i+m)$ statistics. A similar forecasted model was fitted by (Erich J. Plate 2015), Asper his model predicted values, the errors are varied by means of substitution of large number of parameters and also every iteration can be precluded unusual values .However, our predicted model will be very easy and earnestly produced accurate absolute values of minimum rain fall even we substituted larger group of parameters with highest credibility and accuracy.The variance (σ^2) of the model was considered with weighted parameters $\left(\frac{1}{\sigma^2}\right)$ for standardization of data sets because due to floods , cyclonic hit and climatic weather parameter variation. All observations are normally distributed with mean μ and common variance $\left(\frac{1}{\sigma^2}\right)$ *i.e.* We presumed that; all observed values have followed the Gaussianity with weighted values (Weighted logistic regression) $f(x) = N \sim (\mu, \frac{1}{\sigma^2})$. Our predictive run-off errors were produced the best likelihood estimation, it was reduced $\epsilon M_\theta(i+m_j)$ up to 95% (R^2=0.93, Negenkal R square value was 108.8). The present forecasted compartment -model can able to yield good approximate values, even if the data series will follow the heterology distribution. Model iteration has restored at the 100pecent protection @ the time of reproducibility. In the relative comparison of RF (Dharwad district), the maximum distribution of RF was seen in different Progressive months between June (maximum) followed by July (23%) , September (16%) and October (16%) respectively. The results of the predictive rainfall model identified the best fit which has extrapolated various patterns for unobserved components different with larger number of traits (using run off model). The data showed that, the annual maximum daily rainfall will be received @ any specified range between 23.14 mm (minimum) to 83.44 mm (maximum) indicated at large account of fluctuations were seen in the month of Jun and July and also we observed that, maximum rainfall can received with mean temperature (32-33ºC) with moderate precipitation. Log-normal distribution was the best fit probability distribution for majority of the study periods. This model will help us to explore the newer ideas about the prediction of annual maximum daily rainfall to design with small and medium hydraulic pressure. Indirectly we will quantify the soil and water conservation structures, irrigation, drainage works, vegetative waterways and field diversions in the farmers' field.

CONCLUSIONS

Summing of the results concludes that, due to uneven rainfall Agricultural scientists can unable to formulate the cropping patterns in different agro climatic zones at National level. Every year, periodic hitting of cyclones and global warming, the agriculture usual production will be deteorating to achieve sustainable goals. In this programmatic approach, the analytical intervention is very important to elucidate the problem at the early stage. As such being the case, weather and rainfall mathematical /statistical model will be necessary to fill the gap by means of development of new- algorithms of rainfall . An overall essence of the above model we predict the absolute minimum rainfall at the greatest accuracy with reduced errors in the Indian context. The developed runoff-predictive models successfully integrated multiple meteorological and geographic parameters, thereby enhancing the accuracy of rainfall estimation and optimizing decision-making in agriculture. The findings underscore the model's value in forecasting rainfall-dependent agricultural outcomes, managing irrigation resources, and contributing to climate-resilient policy development.

FUTURE SCOPE

Future line of work can be the current model, though focused on Dharwad district, can be extended to other agro-climatic zones in India to generalize the model's applicability. Coupling the predictive models with real-time meteorological data feeds and satellite observations can improve forecasting capabilities. Linking the rainfall prediction model with crop simulation software (e.g., DSSAT, APSIM) can help in optimizing sowing dates and irrigation schedules.

REFERENCES

Beven, K. J. and Freer. J. .E. (2001). Equifinality, data assimilation, and uncertainty estimation inmechanistic modeling of Complex environmental systems using the GLUE Method. *J. Hydrol.*, *249,* 11–29.

Erich, J. Plate (2015). Uncertainty Analysis of Multi-Model Flood Forecasts. *Water*, *7*(12), 6788-6809.

Manikandan, M., Thiyagarajan, G. and Vijaya Kumar, G. (2011). Probability analysis for estimating annual one day maximum rainfall in Tamil Nadu Agricultural University. *J. Madras Agric*., *98*(1-3), 69-73.

Nassif, W. G., Al-Taai, O. T., Mohammed, A. J. and AL-Shamarti, H. K. (2021). Estimate probability distribution of monthly maximum daily rainfall of Iraq *J. Phy. Conf. Ser.*, *5*(3), 1804-1813.

Singh, B., Rajpurohit, D., Vasishth, A. and Singh, J. (2012). Probability analysis for estimation of annual one day maximum rainfall of Jhalarapatan area of Rajasthan India. *Plant Arch.,12*(2): 1093-1100.

How to cite this article: Vinushree R., Ashish Baluni, Megha J. and Chetan Mahadev Rudrapur (2024). Probability Distributions Models to Optimize the Rainfall in Karnataka State. *Biological Forum – An International Journal, 16*(7): 341-347.