Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 4 (2023)

Statistical Analysis and Distribution Modeling of Rainfall Data in Flood-Prone Regions of Karnataka State

Kumudha H R[1] , Kokila Ramesh[2]

[1] *Research Scholar, Jain University, Bangalore and Assistant Professor, Bharathi College PG & RC, Bharathinagara, Mandya, Karnataka, India*

[2] *Assistant Professor, Jain University, Bangalore, India*

Abstract This research paper presents a comprehensive analysis of southwest monsoon rainfall data in flood-prone regions of Karnataka State, employing suitable statistical techniques and data distribution modeling. The detailed analysis of the dataset helped in understanding rainfall patterns, which is essential in identifying flood risks in Karnataka.The analysis includes the computation of descriptive statistics, exploration of data stationarity and the application of distribution models to characterize rainfall patterns. Maximum likelihood estimator has been used to estimate the parameters involved in the data distribution modelling. Further, hypothesis tests were conducted to assess the goodness-of-fit of these distribution models. These models are found to be suitable for the dataset in comparison with the Gaussian models.

keywords:*Flood prone regionsKarnatakaMixture modelsGaussian distribution*

1. Introduction:

Rainfall plays a vital role in the hydrological cycle, but accurately predicting rainfall events at specific locations is challenging due to the intricate processes involved, from convection to cloud formation and precipitation. In India more than 80% of the country's total rainfall occurs during the summer monsoon season which is from June - September.

Rainfall data as time series does not follow a regular pattern which can be shown as non-Gaussian.Most of the studies in the literature assumed rainfall distribution to be Gaussian for the sake of predictions. However, it is essential to acknowledge the non-Gaussian nature of these data, considering various factors affecting the data such as mean, variance, skewness, kurtosis and stationarity. This requires extensive data analysis using suitable techniques.

When dealing with non-Gaussian time series data, it is preferable to define the data in terms of key moments and the power spectral density function. Gaussian processes are commonly used for this purpose, and transformations of Gaussian processes can serve as a non-Gaussian modeling approach. NavneetKumar,BernhardTischbein, MirzaKaleem Begdiscussed a study on rainfall and temperature trends in India's Upper Kharun Catchment. It employs various trend detection methods and introduces a new approach. Rainfall data from 1961 to 2011 showed no significant trends except for an increase in peak monthly rainfall. Temperature data displayed no significant trends, but slight increases in specific months were observed. The study provides valuable insights into local climate dynamics, aiding climate change adaptation and resource management decisions in the region. The study's approach, which combines both parametric and non-parametric tests, including the innovative Gaussian-linear trend detection test, contributes to a more nuanced understanding of climate dynamics in the region. These insights are crucial for informed decision-making, especially in the context of climate change adaptation and resource management.

In the study of Prediction of heavy rainfall days over a peninsular Indian station using the machine learning, the authors Kandula V Subrahmanyam ,Cramsenthiland et al emphasize the significance of predicting heavy rainfall events for effective weather-dependent activities management. They highlight the limitations of traditional Numerical Weather Prediction models and propose the use of machine learning, specifically Gaussian Process Regression (GPR), for predicting heavy and light rainfall days. The study uses 116 years of daily rainfall data from Sriharikota, India, for model training. The GPR model's performance is assessed and compared with other machine learning models such as K-nearest neighbor, random forest, and decision trees. The GPR model exhibits promising results, particularly for heavy rainfall predictions, with low errors (root mean square error = 0.161; mean absolute error = 0.126; mean squared error = 0.026). Furthermore, the GPR model is extended to predict the spatial distribution of monthly rainfall across India. The study suggests that the GPR model could be a valuable tool for predicting heavy rainfall events at specific locations. The authors acknowledge the support from the Indian Space Research Organisation (ISRO) and the use of historical rainfall data from the India Meteorological Department (IMD) for their research.

The study of Best-Fit Probability Models for Maximum Monthly Rainfall in Bangladesh Using Gaussian Distributions, the authors MdAshrafulAlam, Craig Farnham and Kazuo Emura focuses on analyzing extreme precipitation data from 35 weather stations in Bangladesh using various statistical distributions, including Gaussian (normal) distributions and s of multiple Gaussian distributions. The researchers employed maximum likelihood estimation for parameter estimation and used graphical and numerical criteria to determine the best-fit distribution for each station.

The study found that Gaussian (normal) distributions (N) were the best-fit model for 51% of the weather stations of two Gaussian distributions (N2) and three Gaussian distributions (N3) provided the best-fit results for 20% and 14% of the stations, respectively.Five-component Gaussian distribution (N5) was the best-fit for 11% of the stations.The research calculated rainfall heights corresponding to different return periods (10-year, 25-year, 50-year, and 100-year) for each location using the selected distributions.The results have practical implications for policymakers, as they can use this data to plan initiatives aimed at mitigating the impacts of extreme rainfall events, such as floods and landslides, in different regions of Bangladesh.

Overall, this study provides valuable insights into the statistical modeling of extreme precipitation events and their potential consequences, offering a useful tool for risk assessment and disaster management in Bangladesh.

C.A. Glasbey and I.M. Nevison, presented a novel approach to model hourly rainfall data by applying a monotonic transformation to achieve normality. The transformed data create a latent Gaussian variable, allowing for autocorrelation modeling. The study validates the model's performance against real data and traditional point process models, demonstrating its flexibility and potential for disaggregation. This innovative approach has applications in rainfall simulation, forecasting, and fine-resolution data generation. The model's suitability for fitting the Edinburgh rainfall data is comparable to that of well-established point process models.

Niharika Mishra and Ajay Kushwaha, emphasizes the importance of rainfall prediction for water resource management and the challenges posed by dynamic weather patterns. The paper draws upon meteorological data collected by the Department of Agricultural Meteorology at Indira Gandhi Agricultural University, Raipur (C.G.), as the basis for their research. It discusses the application of machine learning, specifically Gaussian Process Regression, to enhance accuracy. Utilizing meteorological data, the study achieves an impressive 95.4% accuracy in rainfall prediction, confirming the practical promise of the proposed model.

Zhengzheng Li, Yan Zhang and Scott E. Giangrande - A study focused on developing a Gaussian rainfall-rate estimator (GMRE) for polar metric radar-based rainfall-rate estimation. The study follows a general framework based on the Gaussian model () and Bayes least squares estimation for weather radar parameter estimations. It highlights the advantages of GMRE, its application across various rain regimes and regions, and its flexibility in incorporating or excluding polar metric radar variables as inputs. GMRE offers several advantages, including its minimum variance unbiased estimation property, adaptability to various conditions, and flexibility in radar variable inputs. The study's evaluation demonstrates GMRE's superior performance over existing techniques, particularly for specific datasets. Future research avenues include combining radar measurements and addressing attenuation issues. GMRE's potential for global applications suggests its significance in advancing accurate rainfall-rate estimation.

PradeebaneVaittinadaAyar, Juliette Blanchet, Emmanuel Paquet and David Penot studied the aims to create a high-resolution spatial rainfall model using station data for hydrological applications, focusing on the Ardèche catchment in southern France. The model combines an autoregressive meta-Gaussian process to account for spatio-temporal dependencies with weather pattern sub-sampling to differentiate rainfall intensity classes. It's designed for mountainous catchments and aims to provide fine-scale precipitation data. The model's novelty lies in this combination. The four-step estimation process involves marginal distribution characterization, parameter mapping, temporal correlation determination, and spatial covariance function establishment. Model evaluation against observations shows strong performance in replicating rainfall statistics with minimal discrepancies. This model offers promise for accurate hydrological modeling in complex terrain.

K'ufre-Mfon E. Ekerete, and et al, focused on understanding and modeling the drop size distribution of rainfall, which is crucial for applications like predicting and mitigating satellite signal attenuation in the millimetre band. Several statistical distributions, including exponential, gamma, and lognormal, have been proposed to model rainfall rates. However, empirical observations have sometimes revealed bimodal distributions. This paper examines how well gamma and lognormal distributions fit empirical rainfall data. In conclusion, this study addresses the challenge of accurately modeling rainfall drop size distributions and their impact on satellite signal attenuation. It highlights the limitations of standard models and introduces alternative models based on Gaussian Models as a step toward improved fitting and a deeper understanding of rainfall distribution patterns.

Moonhyuk Kwon, Hyun-Han Kwon and Dawei Han, introduced a multivariate stochastic soil moisture (SM) estimation approach utilizing a Gaussian- nonstationary hidden Markov model (GM-NHMM) to spatially disaggregate AMSR2 SM data across multiple locations in South Korea's Yongdam dam watershed. In this modeling framework, rainfall and air temperature are included as additional predictors. The GM-NHMM model consists of six states, with three predictors representing an unobserved state linked to SM. The key findings such as rainfall is found to significantly contribute to overall predictability in the GM-NHMM model. It plays a crucial role in estimating local SM not captured by the AMSR2 data. Larger-Scale Dynamics: The AMSR2 data, which captures larger scale dynamic features, aids in identifying regional spatial patterns of SM. It complements the local information provided by weather variables (rainfall and temperature). Comparison with Ordinary Regression Model (OLR): The study compares the efficiency of the GM-NHMM model with that of an ordinary regression model (OLR) using the same predictors. The GM-NHMM exhibits a substantially higher mean correlation coefficient (about 0.78) compared to the OLR (about 0.49). Preservation of Spatial Coherence: The GM-NHMM not only provides a more accurate representation of observed SM but also maintains spatial coherence across all stations reasonably well.

In summary, this study presents a stochastic SM estimation model based on a GM-NHMM to spatially disaggregate AMSR2 SM data across multiple locations. It incorporates rainfall and air temperature as predictors and demonstrates the significance of these variables in improving SM predictability. The GM-NHMM model outperforms the OLR model, offering a better representation of observed SM and preserving spatial coherence effectively.

Amjad Hussein and Safaa K. Kadhem, investigated spatial variation in maximum monthly rainfall in Ireland from 2018 to 2020. Also, study calculates return periods for 50 and 100 years for each station using prediction intervals derived from the posterior predictive distribution of the selected models. This information aids in understanding long-term rainfall rates and planning for mitigating high rainfall risks.It uses Bayesian normal models to identify the best-fitting model for 25 weather stations. Model selection criteria and graphical plots confirm that some stations exhibit greater heterogeneity (three components) while others are more homogeneous (two components). The study also calculates return periods for extreme rainfall events. While effective in revealing data heterogeneity, the method does not consider hidden trends in rainfall rates, which may be explored in future research using hidden Markov models.

S. Ly, C. Charles, and A. Degré, focused on spatial interpolation of daily rainfall data in a hilly region in Belgium, comparing geo-statistical and deterministic methods. The research utilizes 30 years of daily rainfall data from 70 rain gauges in the catchment area, which includes river networks. Key findings and methods include Variogram Modeling, Incorporating Elevation, Interpolation Methods. Several interpolation algorithms, including geostatistical methods (Ordinary Kriging, Universal Kriging, Kriging with an External Drift) and deterministic methods (Inverse Distance Weighting), are compared.

The study uses seven validation rain gauges and cross-validation to assess the performance of these algorithms under different densities of rain gauges. Gaussian model is often the best fit among the semi-variogram models. Interpolation with geostatistical and Inverse Distance Weighting (IDW) algorithms outperforms Thiessen polygon interpolation. Ordinary Kriging (ORK) and IDW are considered the best methods, providing the smallest Root Mean Square Error (RMSE) values in most cases. The choice of interpolation method is crucial when there are very few neighborhood sample points. This research underscores the importance of spatial interpolation in hydrological modeling and demonstrates the effectiveness of geo-statistical methods like Ordinary Kriging and IDW for daily rainfall data in hilly landscapes. Additionally, it provides recommendations for selecting the most suitable interpolation method based on the density of available rain gauge data.

Kumudha H R and Dr. Kokila Ramesh have conducted a study on Indian monsoon rainfall, which presents a detailed review of various models used for modeling and forecasting. Additionally, Kokila Ramesh and Iyengar (2017) introduced an innovative approach employing an ANN model, integrating intra-seasonal and inter-annual variability using the backpropagation algorithm. This novel methodology aimed to model and estimate total precipitation during India's monsoon season. The model features a straightforward architecture comprising 10 input nodes, including 2 nodes each for PRM (Pre-Retreat Monsoon), NEM (Northeast Monsoon), and 6 nodes for SWM (Southwest Monsoon), alongside hidden layers housing five neurons and an output layer. Demonstrating a capability to explain about 94% of the observed inter-annual variability in SWM rainfall data, this model's performance was showcased across four informative subsets spanning the years 1901-2000.

Authors Kumudha H R and Kokila Ramesh have conducted a study on the rainfall data examined in this study at the considered temporal scale demonstrating a highly unstructured pattern. To address this nonlinear and unstructured relationship in the time series, an Artificial Neural Network (ANN) model was employed. This network architecture comprises 15 input nodes representing data from 3 seasons, 5 hidden neurons capturing the intricate nonlinear relationships between current and past seasonal rainfall, and 3 output nodes predicting rainfall for pre-monsoon, monsoon, and post-monsoon seasons. Utilizing 50 years of training data encompassing all three seasons, namely pre-monsoon, monsoon, and post-monsoon, this model harnesses inter-annual and inter-seasonal variability to forecast rainfall across all seasons. The constructed network accounts for 85-95% of the observed seasonal rainfall variance.

The collective review of diverse research articles on rainfall patterns reveals several noticeable research gaps in understanding climate dynamics and predicting extreme weather events. Firstly, the absence of exploration into non-linear trends within rainfall and temperature data limits the complete understanding of climate behavior. These non-linear patterns could hold critical information influencing long-term climate projections and necessitate further investigation. Additionally, while some studies examine into rainfall trends, there's a noticeable insufficient of focus on extreme weather events, particularly in predicting and managing their impacts effectively. Understanding and forecasting such extreme events are essential for designing risk mitigation strategies and adapting to climate change.

Furthermore, while Gaussian models are prevalent in some analysis, the lack of exploration or comparison with other statistical distributions might obstruct the discovery of more accurate models. Introducing Gaussian and Gamma models might address some of these gaps by offering alternative approach to model rainfall.

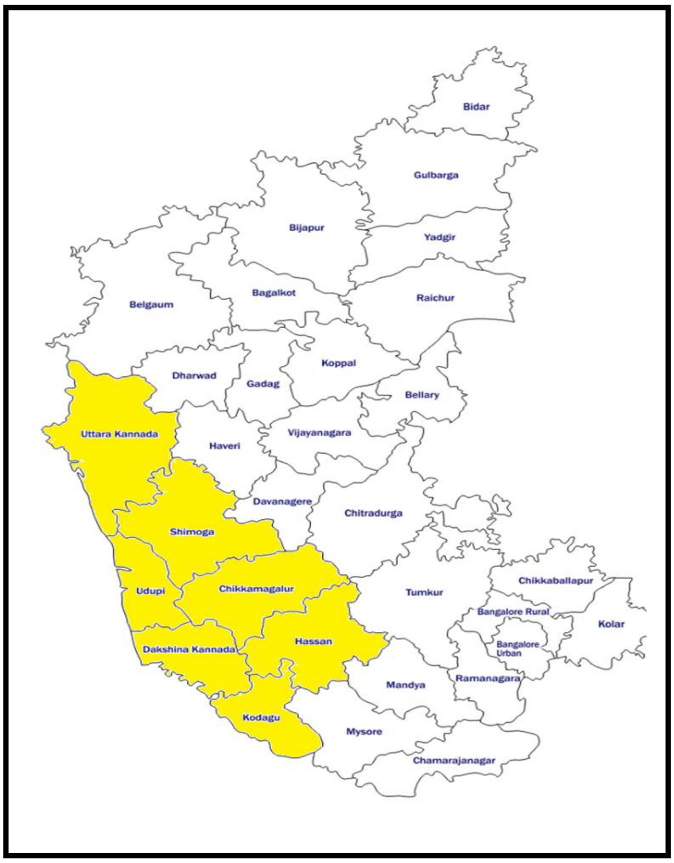
2. DATA:

This paper focuses on modeling flood-prone regions in Karnataka as mentioned in Table 1. These flood prone regions of Karnataka are highlighted in Karnataka map as shown in Figure 1. The study has been restricted to Southwest Monsoon (SWM) rainfall as it contributes 80% of the annual rainfall. The yearly data of SWM rainfall for the period of 51 years (1960- 2010) has been considered for the detailed statistical analysis and modelling. The data has been collected from the website of Indian Institute of Tropical Management (IITM) (http://www.tropmet.res.in) and the Karnataka State Natural Disaster Monitoring Centre (KSNDMC) (https://www.ksndmc.org).The descriptive statistics such as Long term average (LTA), long term deviation (LTD), skewness and kurtosis are tabulated in Table 1.

Table 1: Basic Statistics of SWM Rainfall Data from IITM (1960-2010)

Sub division	LTA (m_R in cm)	LTD (in cm)	Skewness	Kurtosis
Dakshina Kannada	162.78	28.90	0.61	3.52
Udupi	107.58	17.34	0.14	2.73
Uttara Kannada	272.13	36.00	0.36	3.57
Chikkamagalur	102.74	19.58	0.55	3.57
Hassan	49.31	11.59	0.02	2.32
Kodagu	63.61	15.17	0.91	4.26
Shivamoga	99.70	24.31	0.48	3.13

Figure 1: The flood prone regions of Karnataka



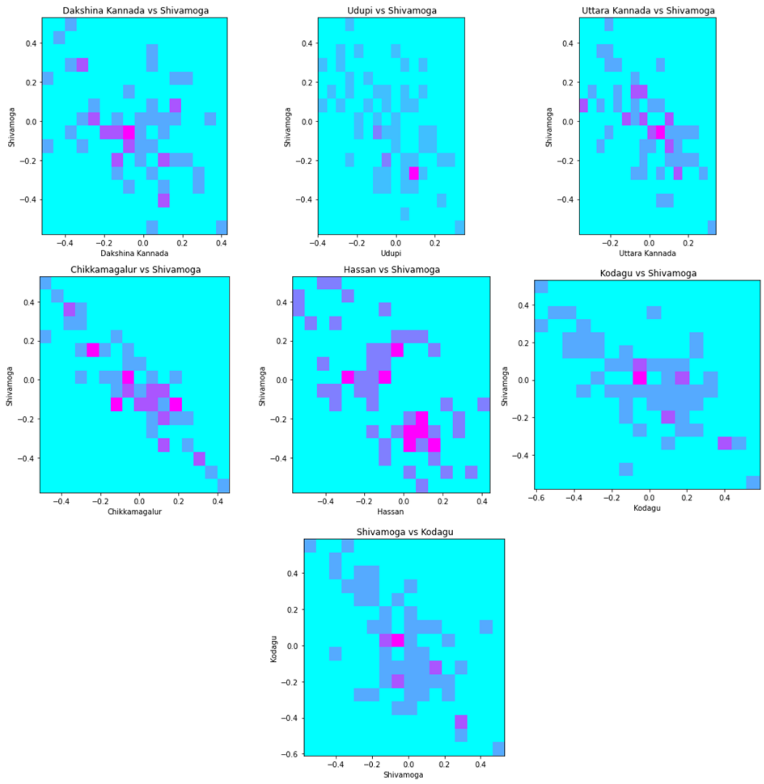
3. DATA ANALYSIS:

To understand the relation between the regions, a heatmap has been plotted and is shown in Figure 2.

This map also gives the relation measure which is known as correlation coefficient for the regions. Since the correlation coefficient is high between the regions for the given sample size, the model developed for one region is relatively suitable to the other flood prone regions.

The heatmap indicates the strength of relationships between different regions in terms of rainfall. Strong positive correlations suggest similar rainfall patterns, while weak correlations imply dissimilar patterns. Dakshina Kannada and Udupi show strong positive correlations with multiple regions, while Chikkamagalur and Hassan exhibit strong correlations with most other regions. Overall, this information helps to understand how rainfall in these regions is interrelated.

Figure 2: Heatmap - To understand the relation between the regions.



Rainfall, considered as a random variable with positive valuesdenoted as R_i, where i ranges from 1 , 2 … 51 for the above mentioned regions are normalized using their respectiveLTA (m_R),

$$Z_i = log\left(\frac{R_i}{m_R}\right)(1)$$

The basic statistics such as mean (m_z), standard deviation(σ_z), skewness (S_z) and kurtosis (k_z) for the normalized data are tabulated in Table 2.

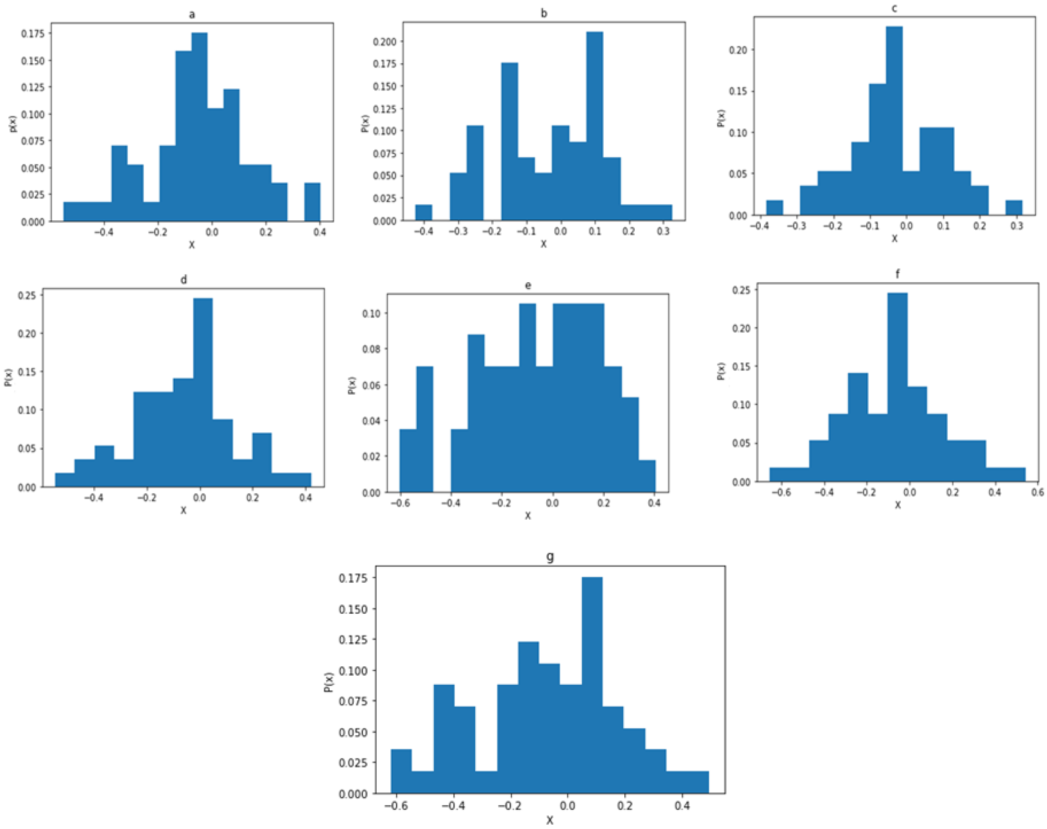
The histogram helps in understanding the distribution pattern of the dataset. The Gaussian distribution ischaracterized by a bell-shaped curve, often representing continuous data and is useful for modeling. By analysing the histogram (Figure 3), it is identified that the distribution characteristics of the southwest monsoon data in the flood prone regions of Karnataka. The histogram displays a bimodal pattern, suggesting the presence of two distinct peaks in the data distribution. This bimodal behavior aligns with the characteristics of a Gaussian curve. Introducing Gaussian and Gamma distribution models could enhance the ability to model more accurately. These models could potentially provide a better fit to the observed data, understanding the nature of monsoon patterns in the region. Integrating these models could lead to a more precise fit to observed data, providing a deeper comprehension of monsoon behavior in the considered regions.

Table 2: Basic Statistics of SWM Rainfall Normalized Data

Sub division	LTA (m_z)	LTD (σ_z)	Skewness	Kurtosis	ADF Statistics
Dakshina Kannada	-0.02	0.18	0.06	3.08	-1.90
Udupi	-0.01	0.16	-0.24	2.58	-5.38
Uttara Kannada	-0.01	0.13	-0.11	3.36	-5.00
Chikkamagalur	-0.02	0.19	-0.09	3.32	-3.67
Hassan	-0.03	0.25	-0.44	2.49	-5.86
Kodagu	-0.03	0.23	0.12	3.29	-6.76
Shivamoga	-0.03	0.25	-0.15	2.73	-4.12

Figure 3: Relative frequency histogram offlood prone regions of Karnataka

Description a) Dakshina Kannada, b) Udupi, c) Uttara Kannada, d) Chikkamagalur, e) Hassan, f) Kodagu and g) Shivamoga



The Augmented Dickey Fuller (ADF) test statistic is used to assess whether the time series data is stationary or non-stationary. In this case, the ADF statistic is highly negative, indicating strong evidence against the presence of a unit root (i.e., non-stationary) in the data. The more negative the ADF statistic, the stronger the evidence against non-stationary.

$$ADF\ Statistic = \frac{Z_i - 1}{standrad\ error(Z_i)} \quad (2)$$

The p-value is approximately close to zero. The p-value represents the probability of observing the ADF statistic under the null hypothesis that the data has a unit root (i.e., it is non-stationary). In this case, the extremely low p-value indicates very strong evidence against the null hypothesis. Since the p-value is much smaller than a typical significance level (e.g., 0.05), you would typically reject the null hypothesis, suggesting that the data is stationary. The critical values are thresholds used for comparison with the ADF statistic. These critical values depend on the desired level of significance (e.g., 1%, 5%, or 10%). In your result, the ADF statistic is significantly lower than all of the critical values, further supporting the conclusion that the data is likely stationary. In summary, based on the ADF test result it is provided that, the ADF statistic is highly negative, indicating strong evidence against non-stationary. The extremely low p-value (close to zero) indicates very strong evidence against the null hypothesis of non-stationary. The ADF statistic is significantly lower than the critical values at various significance levels, reinforcing the conclusion that the data is likely stationary. Overall, the statistical tests and measures indicate that while there are some deviations from a perfectly Gaussian distribution (e.g., slight skewness and kurtosis different from 3), these deviations are shown non-Gaussian of the data also the figure 1 shows that the data has bimodal. Therefore a Gaussian and Gamma model have been introduced.

4. DESCRIPTION OF THE GAUSSIAN AND GAMMA MODEL:

In the previous section the data for the study has shown to be non-Gaussian through ADF test hence there is a need to model the data using non-Gaussian structure. The work by Kokila and Iyengar explains that a Gaussian mixture model was necessary for core monsoons and subdivisions regions of India. This situation can be effectively addressed by modeling it as a function of a Gaussian and Gamma Model. An illustrative example is the modeling of the rainfall process, which often demonstrates such characteristics. Therefore, in this paper, an attempt has been undertaken to represent Indian monsoon rainfall for the flood prone regions of Karnataka as a function derived from a Gaussian and Gamma Distribution Model.

Initially, the proposal involves suggesting a combination of two Gaussian random variables, denoted as xand y, with a certain proportion α for the transformed data z. The subsequent equation offers an illustration of how these Gaussian random variables x andy, which are independently and identically distributed, are blended using the proportion w_i.

Let $z = ux + (1-u)y$ (3)

Such that $m_{z|u} = um_x + (1-u)m_y$ and $\sigma_z^2 = u^2\sigma_x^2 + (1-u)^2\sigma_y^2$ (4)

In the context of a Bernoulli random variable, it's common to describe it in terms of a probability mass function (PMF) rather than a probability density function (PDF). The PMF of a Bernoulli random variable is typically used to specify the probabilities of discrete outcomes.

A Bernoulli random variable, denoted as u, has two possible outcomes: 0 and 1, with probabilities p and $1-p$, respectively.

$$p(u) = w\delta(u-1) + (1-w)\delta(u-0) \quad (5)$$

The conditional density function of z given u is as follows

$$f(x{:}\,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]} \text{and}$$

$$f(y{:}\,\alpha,\beta) = \frac{z^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} (6)$$

Consider,

$$p(z) = \frac{w_i}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]} + (1-w_i)\frac{z^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} (7)$$

Here the parameters μ, σ^2, α and β are found using maximum likelihood (MLH) estimator method . The parameters μ, σ^2, α and β are the moments of the Gaussian and gamma random variables x and y respectively. The MLH function L of the above equation is given by

$$L(\mu, \sigma^2, \alpha, \beta) = \prod_{i=1}^{n} [w_i * f_{gaussian}\ (x: \mu, \sigma^2) + (1 - w_i) * f_{gamma}\ (y: \alpha, \beta)] \qquad (8)$$

Instead of maximizing this product, which can be quite laborious, often take advantage of the fact that the logarithm is a monotonically increasing function. Consequently, maximizing the log-likelihood is considered equivalent and is a more convenient approach.

$$\log[L(\mu, \sigma^2, \alpha, \beta)] = \sum_{i=1}^{n} \log[p(z_i | \mu, \sigma^2, \alpha, \beta)] \qquad (9)$$

By taking partial differentiations of the above equation (9) with respect to the parameters, one at a time, and setting them equal to zero, then derive the parameter expressions. The parameters values are tabulate in Table 3.

5. GAUSSIAN AND GAMMA MODEL ON RAINFALL DATA

Table 3: The parameter values of equation (7) for flood prone regions of Karnataka

Region	w	μ_R	σ	α	β
Dakshina Kannada	0.3	-0.02	0.15	1055.04	-5.6
Udupi	0.6	-0.01	0.10	403.8	-3.31
Uttara Kannada	0.48	-0.00	0.12	573.74	-3.19
Chikkamagalur	0.35	-0.02	0.17	978.34	-5.91
Hassan	0.45	-0.02	0.14	313.44	-4.45
Kodagu	0.37	-0.03	0.21	376.11	-4.46
Shivamoga	0.41	-0.03	0.18	489.21	-5.44

Equation (7) has been confirmed to meet all the criteria for being a valid probability density function. This model is then applied to the original data using the transformation described in equation (10). The resulting probability density function for R can be expressed as follows:

$$p(R) = \frac{1}{R}\left[\frac{w_i}{\sigma\sqrt{2\pi}} e^{\left[-\frac{\left(\log\left(\frac{R}{\mu_R}\right)-\mu\right)^2}{2\sigma^2}\right]} + (1 - w_i)\frac{\left(\log\left(\frac{R}{\mu_R}\right)\right)^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\right] \quad R>0 \qquad (10)$$

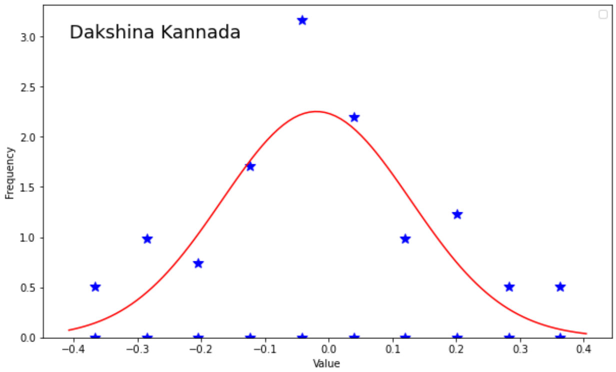
Figure 4 displays a comparison between the actual data and that of the Gaussian Mixture distribution model and Figure 5displays a comparison between the actual data and that of the Gaussian and Gamma distribution model. It is visually evident that, the Gaussian and Gamma distribution closely adheres to a probability density function as defined in equation (10). Table 4 displays the first four moments of the Gaussian and Gamma modelwhich are then compared with the moments of the actual data.

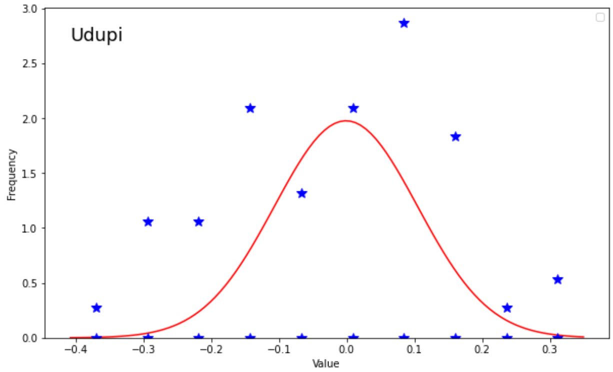
In the event that assume the data follows a Gaussian and Gamma distribution, the skewness and kurtosis of the data should ideally be 0 and 3. Hence, this model is deemed appropriate for the utilized dataset.Given that both models yielded favourable results, signifying skewness close to 0 and kurtosis around 3, the Gaussian and gamma models stand as potential considerations for subsequent analyses.

Table 4: Comparison between the actual data moments with the model data moments

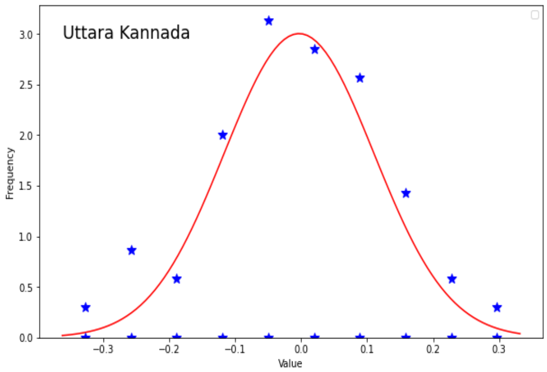
Region	Actual Data Moments				Model Data Moments			
	μ_a (in cm)	σ_a (in cm)	S_a	K_a	μ_m (in cm)	σ_m (in cm)	S_m	K_m
Dakshina Kannada	162.78	28.9	0.61	3.52	160.35	28.18	0.31	3.22
Udupi	107.58	17.34	0.14	2.73	106.19	16.94	0.14	2.83
Uttara Kannada	272.13	36.01	0.36	3.57	270.69	35.58	0.31	3.57
Chikkamagalur	102.74	19.58	0.55	3.57	100.94	19.05	0.02	3.57
Hassan	49.31	11.59	0.02	2.32	47.91	11.15	0.10	2.92
Kodagu	63.61	15.17	0.91	4.26	61.94	14.62	0.90	3.26
Shivamoga	99.70	24.31	0.48	3.13	96.83	23.38	0.48	3.13

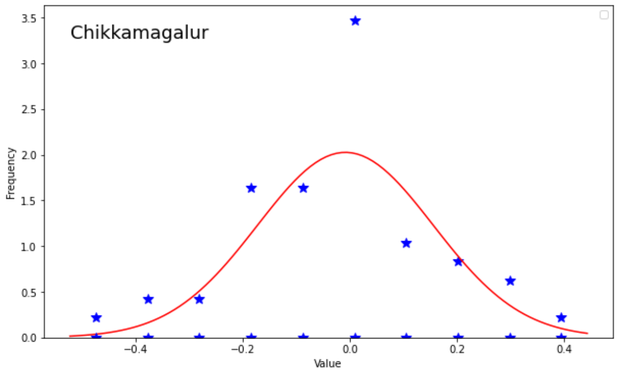
Figure 4:Comparison between the actual data and the model data using Gaussian Mixture Model

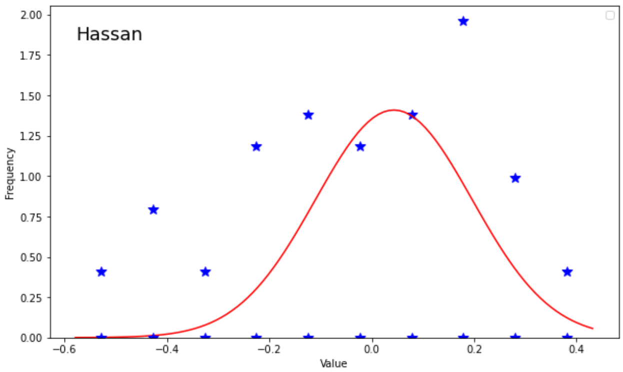


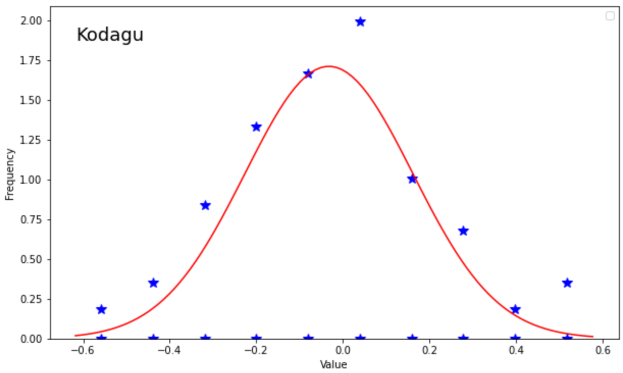


Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 4 (2023)









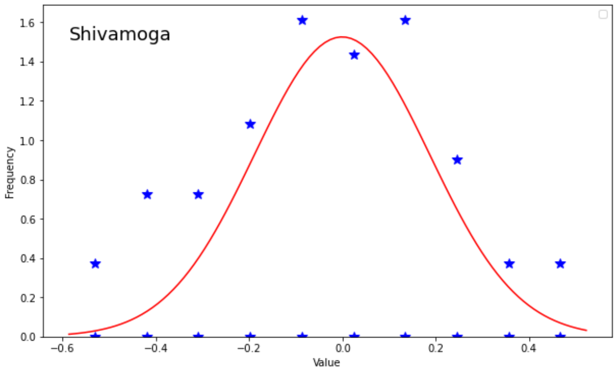
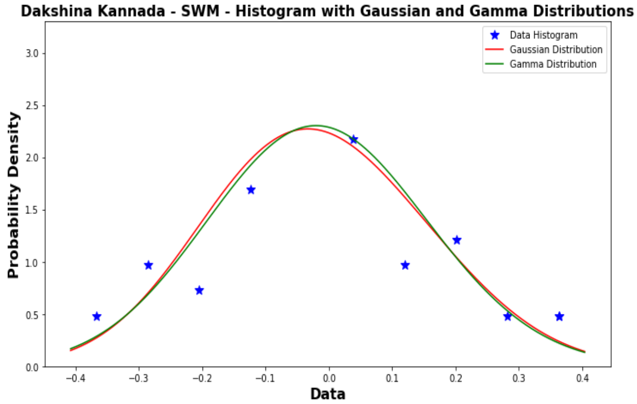
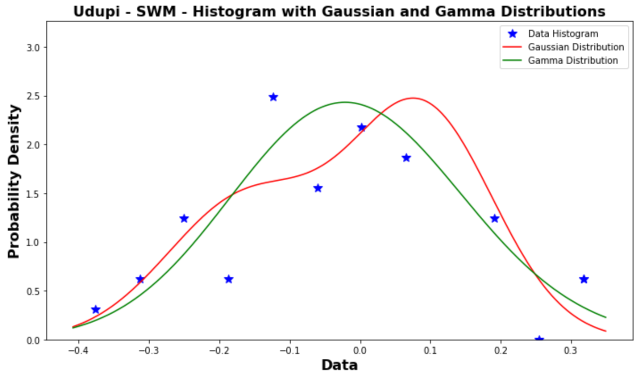
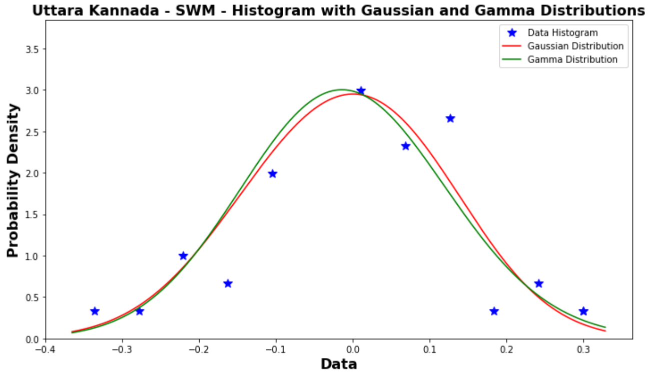
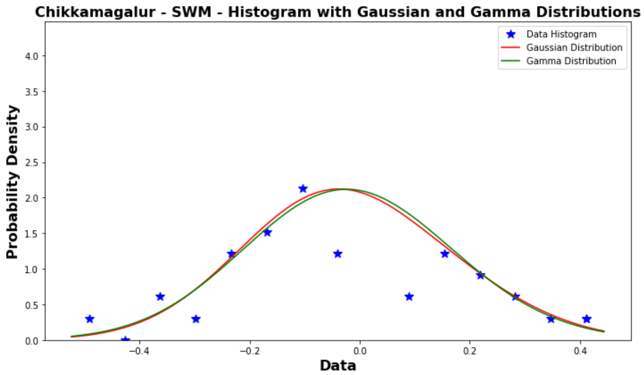


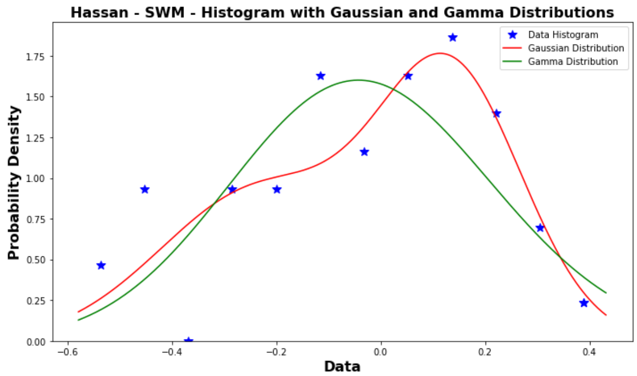
Figure 5: Comparison between the actual data and the model data using Gaussian and Gamma Model

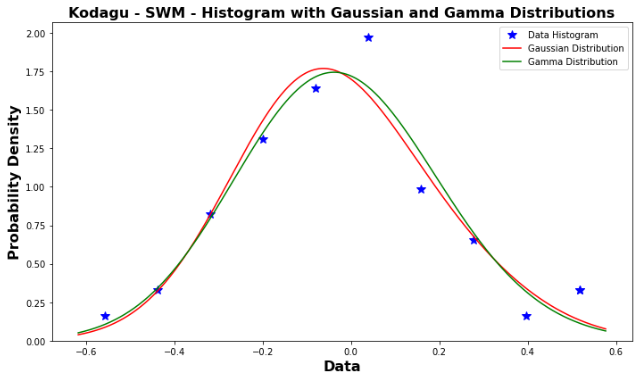


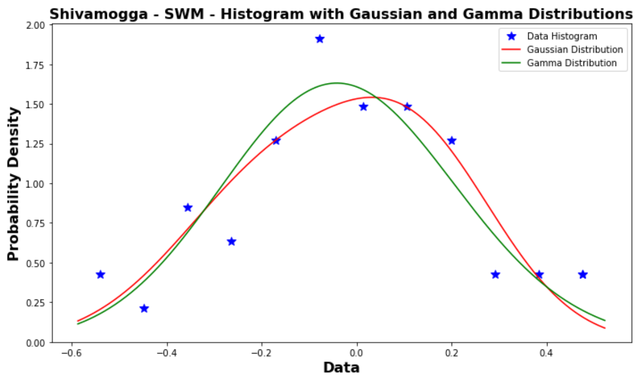












6. Discussion And Conclusion:

In this research, a comprehensive analysis of a dataset using statistical methods and distribution modeling is conducted. The objective was to gain insights into the underlying data structure and assess the goodness-of-fit of specific probability distributions. It is began by calculating fundamental statistics, including the mean, standard deviation, skewness, and kurtosis, which provided key information about the data's central tendency, dispersion, and shape. To facilitate the analysis, data normalization techniques were employed, particularly taking the logarithm of the data relative to its mean. This step not only aided in making the data conform more closely to a normal distribution but also enabled us to perform the Augmented Dickey-Fuller test for stationary. The results of this test were crucial in determining whether the data exhibited temporal patterns.

Furthermore,distribution modeling techniques were employed including a Gaussian Model and gamma distribution, to characterize the data's probability distribution. The model allowed us to represent the data as a combination of multiple Gaussian distributions, while the gamma distribution was chosen to model the skewness often observed in positive-valued datasets. Visualizations were created to provide a clear representation of the probability density functions of these models and compare them to observed data. Hypothesis testing, including the Chi-Square and Anderson-Darling tests, was performed to assess the goodness-of-fit of these models and determine if the data followed a specific distribution. The findings contribute valuable insights into the statistical properties of the dataset and lay the groundwork for further analysis and modeling.

The research conducted by Kokila Ramesh and R N Iyengar examines a non-Gaussian model for Indian monsoon rainfall. Their initial work focused on employing a Gaussian Mixture Model to study Indian monsoon rainfall. The authors used the Gaussian Mixture Model before is known to be really good at understanding complicated data patterns. In this paper, the authors expanded the study by introducing and utilizing Gaussian and Gamma distribution models to explore and analyzethe rainfall patterns specifically in areas like Dakshina Kannada, Udupi, Uttara Kannada, Chikkamagalur, Hassan, Kodagu, and Shivamoga.

However, in this study, the introduction and exploration of the Gamma distribution model alongside the Gaussian model. Through the analysis, both models showcase promising results, emphasizing the suitability and effectiveness of these models help us figure out and understand the complex details of how rainfall pattern behaves during the monsoon season. By demonstrating that both models exhibit favorable outcomes in analyzing the southwest monsoon rainfall patterns.

In conclusion, the Gaussian and Gamma distribution models in the study of southwest monsoon rainfall patterns provide a valuable contribution to the field. The successful application of these models underscores their potential as reliable tools for meteorological analysis. This study suggests that the Gaussian and Gamma distribution models can complement the existing Gaussian Mixture Model, offering an enhanced understanding of rainfall variability and aiding in more accurate predictions and assessments. Continued research in this direction could lead to benefitting forecasts and climate studies.

References:

1. Kumar, N., Tischbein, B., & Beg, M. K. (2019). Multiple trend analysis of rainfall and temperature for a monsoon-dominated catchment in India. *Meteorology and Atmospheric Physics*, 131(5), 1-15.
2. Subrahmanyam, K. V., Cramsenthil, & et al. (2021). Prediction of heavy rainfall days over a peninsular Indian station using machine learning algorithms. *Journal of Earth System Science*, *130(240). Indian Academy of Sciences*.
3. Alam, M. A., Farnham, C., &Emura, K. (2018). Best-Fit Probability Models for Maximum Monthly Rainfall in Bangladesh Using Gaussian Distributions. *Geosciences*, 8(4), 138.
4. Glasbey, C. A., &Nevison, I. M. (1997). Rainfall Modeling Using a Latent Gaussian Variable. In Modelling Longitudinal and Spatially Correlated Data (pp. 233-242). *United Kingdom: Vol 122*.
5. Mishra, N., &Kushwaha, A. (2019). Rainfall Prediction using Gaussian Process Regression Classifier. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 8(8), ISSN: 2278 – 1323.
6. Li, Z., Zhang, Y., &Giangrande, S. E. (2012). Rainfall-Rate Estimation Using Gaussian Parameter Estimator: Training and Validation. *Journal of Atmospheric and Oceanic Technology*, 29(5), 01 May 2012.
7. Ayar, P. V., Blanchet, J., Paquet, E., &Penot, D. (2020). Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model. *Journal of Hydrology*, 581, 124451.

8. Ekerete, K. M., Hunt, F. H., Jeffery, J. L., &Otung, I. E. (2015). Modeling rainfall drop size distribution in southern England using a Gaussian Model. *Radio Science*, 50(9), 853-957.
9. Kwon, M., Kwon, H. H., & Han, D. (2018). A spatial downscaling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian- nonstationary hidden Markov model. Journal of Hydrology, 564, 1194-1207.
10. Hussein, A., &Kadhem, S. K. (2022). Spatialmodeling for analyzing a rainfall pattern: *A case study in Ireland. Open Engineering*, 12, 204-214. De Gruyter.
11. Ly, S., Charles, C., &Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. HESS, 15(7), 2259-2274. *European Geosciences Union - Hydrology and Earth System Sciences*.
12. Kokila Ramesh and R.N.Iyengar (2017). A non-Gaussian model for Indian monsoon rainfall, *International Journal of research – Granthaalayah*. Vol.5 (Iss.4: RAST), ISSN- 2350-0530(O), ISSN- 2394-3629(P).
13. Kokila Ramesh, & Kumudha, H. R. (2019). A review on forecasting Indian monsoon rainfall, *International Journal of Innovative Science and Research Technology*, Special Issue(AAM 2019), 9– 14.
14. Kokila Ramesh, & Iyengar, R. N. (2017). Forecasting Indian monsoon rainfall including within year seasonal variability. *International Journal of Civil Engineering and Technology (IJCIET*), 8(2), 390–399.
15. Kumudha H. R,& Kokila Ramesh. (2023). Forecasting of Karnataka Seasons Rainfall Data Using ANN Approach. *Journal of Survey in Fisheries Sciences*, 10(1S) 3431-3448.