ISSN: 2645-8489
Original Article

Modified Extended Kumaraswamy Exponential Type-I Distribution on Environmental data

Umesh Kumar Yadav[1], Abhay Kumar[2]

[1] Mahendra Bindeswori Multiple Campus, Rajbiraj
Tribhuvan University
Email: yadavumesh398@gmail.com
[2] Government Engineering College Samastipur, India
Email: abhay426@gmail.com

Corresponding Author
Madan Umesh Kumar Yadav
Email: yadavumesh398@gmail.com

Funding: This research received no specific grant from any funding agency in the Public, commercial, or not-for-profit sectors.

Abstract

This paper introduces a novel probability distribution, termed the Modified Extended Kumaraswamy Exponential Type-I Distribution (MEKWE), formulated by modifying the New Extended Kumaraswamy Exponential Distribution (NEKWE). The MEKWE distribution is characterized by its flexibility and ability to model data exhibiting various shapes such as skewness and kurtosis. The performance of the MEKWE distribution is evaluated using Maximum Likelihood Estimation (MLE) for parameter estimation and validated using real-world environmental data from Nepal (TSP of 2023). The distribution's fit is assessed through various statistical tests, including the Kolmogrov-Smirnov (KS) test, AIC, BIC, and graphical diagnostics like histograms, box plots, and TTT plots. The MEKWE distribution outperforms other candidate distributions in terms of fitting the environmental data, demonstrating its potential application in environmental modeling, reliability engineering, and other fields requiring flexible modeling of lifetime data.

Keywords: *Environmental Data, Kolmogrov-Smirnov Test, Kumaraswamy Exponential Distribution, Maximum Likelihood Estimation, Skewness, Kurtosis.*

JEL Classification: *C02, C10, C15, Q50, Q53, Q58*

Introduction

This manuscript presents the preliminary findings of an ongoing study. The current work focuses on "analysis of Air Quality dataset of Kathmandu using Probability Distribution", and while the broader research project is still in progress, these results provide an early insight into key objectives and findings for upcoming research. Further analyses and additional data will be included in subsequent publications to complete the research."

Kathmandu, the capital of Nepal, is a vibrant city nestled in a bowl-shaped valley surrounded by hills at an altitude of approximately 1,400 meters above sea level. While its historic and cultural heritage makes it a key tourist

How to cite this article (APA): Yadav, U. K., & Kumar, A. (2024). Modified Extended Kumaraswamy Exponential Type-I Distribution on Environmental data. *DEPAN, 6*(1), 88-94.

destination, Kathmandu faces significant environmental challenges due to its rapid urbanization, population growth, and topographical constraints. Over the past few decades, the city has witnessed an alarming rise in air pollution levels, exacerbated by vehicular emissions, construction activities, industrial operations, and the burning of biomass for cooking and heating.

Air pollution has become one of Kathmandu's most pressing environmental and public health concerns. Key pollutants such as PM2.5 (fine particulate matter), PM10 (coarse particulate matter), and Total Suspended Particulates (TSP) are critical indicators of air quality. These particulates, often originating from dust, emissions, and combustion processes, have severe implications for human health, contributing to respiratory diseases, cardiovascular issues, and reduced life expectancy. The World Health Organization (WHO) and national guidelines have consistently highlighted that air pollution levels in Kathmandu frequently exceed safe limits, particularly during the winter months.

The valley's unique topography and meteorological conditions further exacerbate pollution levels. During winter, temperature inversions trap pollutants close to the ground, leading to poor air quality. Conversely, the monsoon season typically brings improved conditions as heavy rainfall helps to wash out suspended particulates from the atmosphere. These seasonal variations underscore the importance of studying climatic factors, including maximum temperature (Tem_max), minimum temperature (Tem_min), and precipitation, which play a crucial role in influencing pollutant dispersion and accumulation. This paper introduces a novel probability distribution, termed the Modified Extended Kumaraswamy Exponential Type-I Distribution (MEKWE), formulated by modifying the New Extended Kumaraswamy Exponential Distribution (NEKWE) given by Chaudhary et al. (2024). To test the applicability of the proposed model, model is applied on total suspended particles (TSP) of environmental data of Kathmandu valley.

Literature Review

The Kumaraswamy distribution, first proposed by Kumaraswamy (1980), is widely recognized for its simplicity and flexibility in modeling various distributions. Its general form allows for the modeling of skewed data, making it suitable for diverse applications such as hydrology, engineering, and environmental science. The Exponential distribution has long been used in reliability analysis due to its memoryless property, particularly in modeling the time until an event occurs.

In recent years, the Kumaraswamy Exponential distribution (KE) has gained attention for its ability to model data that exhibit both exponential and Kumaraswamy-like behaviors. The New Extended Kumaraswamy Exponential Distribution (NEKWE), introduced by Chaudhary et al. (2024), extends the KE distribution by incorporating additional parameters to account for various shapes in the data distribution, including heavier tails and different skewness characteristics.

Several works have extended the classical Kumaraswamy and Exponential distributions to improve their applicability in modeling more complex real-world phenomena. For example, the Generalized Kumaraswamy distribution and Extended Exponential distributions have been proposed to model lifetime data with varying hazard rates and aging patterns (see Salazar et al., 2019; Kundu et al., 2021).

Various modification and generalization of exponential are used to explain the survival data. Some of the generated models are generalized exponential proposed by Gupta and Kundu in 2001, two-sided generalized exponential generated by Korkmaz et al. in 2015. Similarly, Telee & kumar introduced modified generalized exponential distribution in 2023. The beta exponential (Nadarajah & Kotz, 2006) is also the modified distribution using exponential distribution. Cordeiro and de Castro (2011) introduced Kumaraswamy exponential model. The gamma exponentiated exponential generated by Ristic & Balakrishnan, 2012) as well as Kumaraswamy transmuted exponential introduced by Afriya et al. in 2016 are generalization of exponential model. Chaudhary et al. (2024) also analyzed the air quality of Kathmandu using a probability model named the New Extended Kumaraswamy

The Modified Extended Kumaraswamy Exponential Type-I Distribution (MEKWE) proposed in this study builds on these developments by incorporating a more flexible form of the distribution that accommodates additional variations in skewness and kurtosis. This modification significantly improves the fit of the distribution to real-world data.

Methodology

1. **Formulation of the MEKWE Distribution:** The probability density function (PDF) and cumulative distribution function (CDF) of the MEKWE distribution wii be defined.
2. **Parameter Estimation:** The parameters α, β, and λ of the MEKWE distribution are estimated using the Maximum Likelihood Estimation (MLE) method, which maximizes the likelihood function based on the observed data.
3. **Model Fitting and Statistical Tests:**
   - The Kolmogrov-Smirnov (KS) test is applied to assess the goodness of fit for the MEKWE distribution. The p-value of the KS test helps determine if the distribution is a good fit for the data.
   - AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are used for model selection, comparing the MEKWE distribution with other competing distributions (e.g., Exponential, Kumaraswamy).
   - Graphical diagnostics, including histograms, box plots, and TTT (Total Time on Test) plots, are employed to visually inspect the fit of the distribution to the data.
4. **Data Analysis:** The MEKWE distribution is tested on a dataset of 2023 environmental data from Kathmandu, Nepal. Descriptive statistics such as mean, skewness, and kurtosis are calculated to summarize the data's characteristics. The model is then fitted to this data, and the fit is assessed using the aforementioned statistical methods.

Model Formulation

The CDF and PDF of the New Extended Kumaraswamy Exponential Distribution (NEKWE) distribution is given as

$$F(x;\alpha,\beta,\lambda)=\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x^{\alpha}}\right)\right\}\right]^{\beta} ;x\geq 0,\alpha,\beta,\lambda>0 \tag{1}$$

Corresponding PDF

$$f(x;\alpha,\beta,\lambda)=\lambda\beta\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x^{\alpha}}\right)\right\}\right]^{\beta-1}\exp\left\{-\lambda x\left(1-e^{-\lambda x^{\alpha}}\right)\right\}\left[1-e^{-\lambda x^{\alpha}}\left(1-\alpha\lambda x^{\alpha}\right)\right] \tag{2}$$

Given cdf in (1) is modified by taking inverse of x. (3) and (4) give the CDF and PDF of the proposed model NEKWE

$$F(x;\alpha,\beta,\lambda)=\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x e^{\alpha/x}}\right)\right\}\right]^{\beta} ;x\geq 0,\alpha,\beta,\lambda>0 \tag{3}$$

Corresponding pdf of the model is given by

$$f(x;\alpha,\beta,\lambda)=\beta\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x e^{\alpha/x}}\right)\right\}\right]^{\beta-1}\exp\left\{-\lambda x\left(1-e^{-\lambda x e^{\alpha/x}}\right)\right\}$$

$$\lambda\left[\left(1-e^{-\lambda x e^{\alpha/x}}\right)+x\lambda e^{-\lambda x e^{\alpha/x}}e^{\alpha/x}\left(1+x^{2}\alpha\right)\right];x\geq 0,\alpha,\beta,\lambda>0 \tag{4}$$

Survival function/ Reliable function:

The survival function of the model proposed here is given by

S(x) = 1 – F(x)

$$=1- F(x;\alpha,\beta,\lambda)=1-\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x^{\alpha}}\right)\right\}\right]^{\beta} \quad ; x\geq 0, \alpha, \beta, \lambda >0$$

Hazard rate function:

The hazard rate function of the model is given by
h(x)= f(x)/(1-F(x))=

$$\beta\left[1-\exp\left\{-\lambda x\left(1-e^{-\lambda x e^{\alpha/x}}\right)\right\}\right]^{-1} \exp\left\{-\lambda x\left(1-e^{-\lambda x e^{\alpha/x}}\right)\right\}$$

$$\lambda\left[\left(1-e^{-\lambda x e^{\alpha/x}}\right)+x\lambda e^{-\lambda x e^{\alpha/x}} e^{\alpha/x}\left(1+x^{2}\alpha\right)\right]; x\geq 0, \alpha, \beta, \lambda >0 \qquad (5)$$

The cdf and pdf curves for the proposed model are shown in figure1

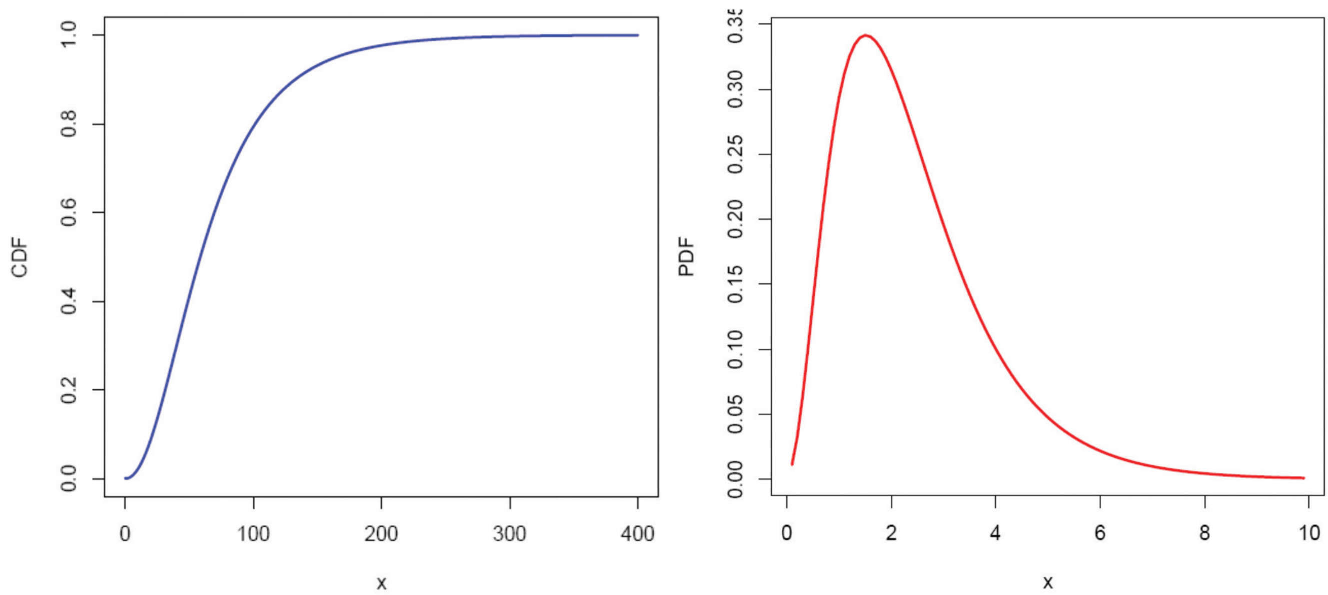


Figure 1: CDF curve (Left) and PDF curve (Right) of MEKWE

Parameter Estimation:

Parameters of the model are estimated using Maximum Likelihood Estimation (MLE). The likelihood function of the MEKWE is given as

$$l(x;\alpha,\beta,\lambda)=n\log\beta+\left(\beta-1\right)\sum_{i=1}^{n}\log\left[1-\exp\left\{-\lambda x_i\left(1-e^{-\lambda x_i e^{\alpha/x_i}}\right)\right\}\right]+$$

$$\sum_{i=1}^{n}\left\{-\lambda x_i\left(1-e^{-\lambda x_i e^{\alpha/x_i}}\right)\right\}+\lambda\sum_{i=1}^{n}\log\left[\left(1-e^{-\lambda x_i e^{\alpha/x_i}}\right)+x_i\lambda e^{-\lambda x_i e^{\alpha/x_i}} e^{\alpha/x_i}\left(1+x_i^{2}\alpha\right)\right]; \qquad (6)$$

First order and second order partial derivatives are obtained with respect to estimate the unknown parameter to estimate the parameters. It is quite difficult to estimate the parameters analytically so we have used optim () function (R Core Team, 2020) of R programming language to estimate the parameters.
This section represents the analysis of real data set to verify the proposed model. The data set is the TSP value of Kathmandu valley provided by Environment department for 2023.
Summary Statistics of the data set is given in table 1.

Table 1

Summary Statistics

Min.	Q1	Median	Mean	Q3	Max	Skewness	Kurtosis
9.088	109.409	154.133	163.642	213.977	371.614	0.3393403	2.636197

Estimated parameters AIC and BIC of the distribution for the TSP are tabulated in Table2.

Table 2

Estimated parameters, AIC and BIC

Alpha	Beta	Lambda	AIC	BIC
53.73	3.91	0.0128	3248.53	3259.467

To know the nature of data, TTT plot and boxplot are plotted and shown I figure 2.

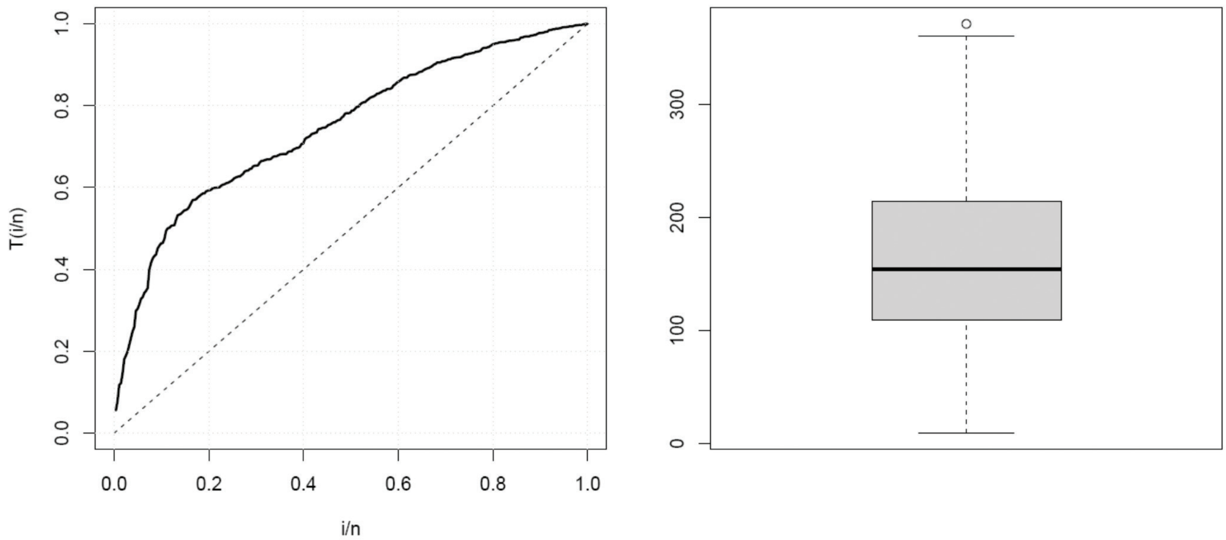
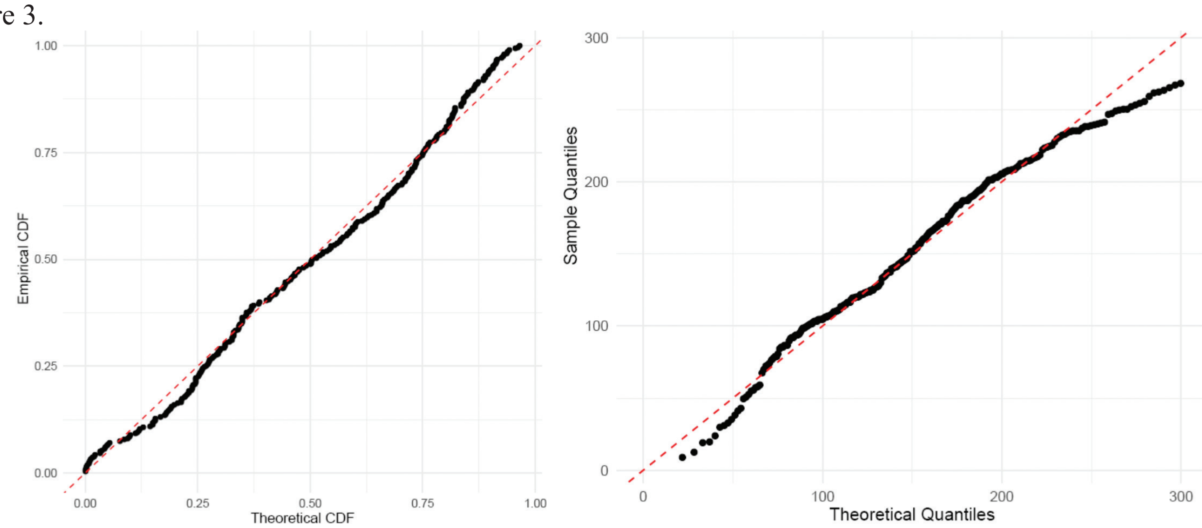


Figure 2: TTT plot (Left) and Boxplot (Right)

To show the model validation on given set of data, PP plot and QQ plot for the model are plotted and shown in figure 3.



One-sample Kolmogrov-Smirnov test is performed to test the goodness for fit the data

Data: D = 0.052711, p-value = 0.4113

Histogram versus fitted pdf for the model are plotted and displayed in figure 4

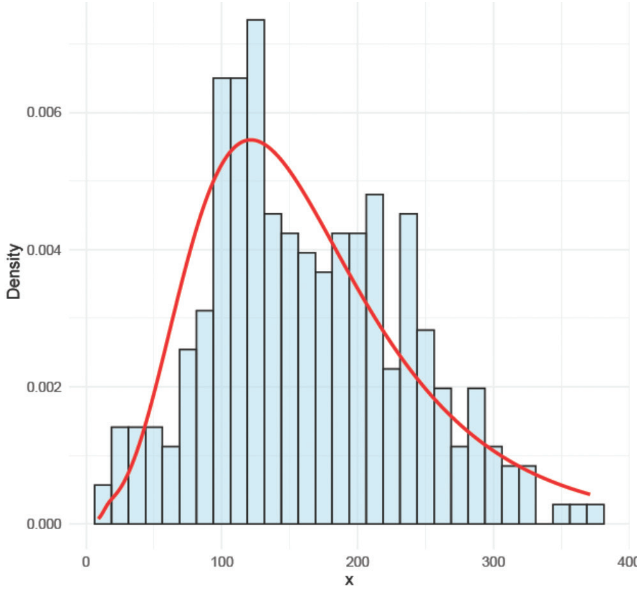


Figure 4: Histogram versus fitted PDF

Results

Following main results are obtained on analysis

- Kolmogorov-Smirnov Test: The results from the one-sample KS test showed that the MEKWE distribution provided a good fit to the environmental data with a p-value of 0.4113, suggesting no significant deviation from the theoretical distribution.
- Skewness and Kurtosis: The data had a skewness of 0.3393 and kurtosis of 2.6362, indicating slight skewness and near-normal peakedness, which is well captured by the MEKWE distribution.
- Descriptive Statistics: The summary statistics for the data indicate a wide range of values, with a minimum of 9.088 and a maximum of 371.614, showing the variability in the environmental measurements.
- AIC and BIC Values: The AIC of the MEKWE model is 3248.53, and the BIC is 3259.467, which indicates that the MEKWE distribution is a competitive model for the data.

Conclusion

These results represent an initial part of a larger ongoing study. Future work will expand on these findings and include additional data to provide a more comprehensive understanding of "Analysis of Air Quality dataset of Kathmandu using Probability Distribution".

The Modified Extended Kumaraswamy Exponential Type-I Distribution (MEKWE) successfully captures the underlying distribution of environmental data from Nepal, demonstrating its effectiveness in modeling data with skewed and kurtotic characteristics. The distribution offers greater flexibility compared to standard distributions like the Exponential and Kumaraswamy, particularly for modeling complex lifetime data with varying hazard rates. The performance of the MEKWE distribution, evaluated using MLE, KS tests, AIC, and BIC, indicates that it is a reliable model for a wide range of applications, including environmental science, reliability engineering, and other fields requiring sophisticated statistical modeling. Future work may focus on extending the MEKWE distribution to handle censored data and incorporating additional features such as time-varying parameters.

References

Afifya, A. Z., Cordeiro, G. M., Yousof, H. M., Nofal, Z. M., & Alzaatreh, A. (2016). The Kumaraswamy transmuted-G family of distributions: properties and applications. *Journal of Data Science, 14*(2), 245-270.

Almongy, H. M., Almetwally, E. M., Alharbi, R., Alnagar, D., Hafez, E. H., & Mohie El-Din, M. M. Chaudhary, A. K., Telee, L. B. S., Karki, M., & Kumar, V. (2024). Statistical analysis of air quality dataset of Kathmandu, Nepal, with a New Extended Kumaraswamy Exponential Distribution. *Environmental Science and Pollution Research,* 1-16.

Cordeiro, G. M., & de Castro, M. (2011). A new family of generalized distributions. *Journal of statistical computation and simulation, 81*(7), 883-898. https://doi.org/10.1080/00949650903530745

Gupta, R. D., & Kundu, D. (2001). Generalized exponential distribution: different methods of estimations. *Journal of Statistical Computation and Simulation, 69*(4), 315-337.

Korkmaz, M. Ç., & Genç, A. İ. (2015). Two-sided generalized exponential distribution. *Communications in Statistics-Theory and Methods, 44*(23), 5049-5070.

Nadarajah, S., & Kotz, S. (2006). The beta exponential distribution. *Reliability engineering & system safety, 91*(6), 689-697. https://doi.org/10.1016/j.ress.2005.05.008

Ristic, M. M., & Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of statistical computation and simulation, 82*(8), 1191-1206. https://doi.org/10.1080/00949655.2011.574633

Telee, L. B. S., & Kumar, V. (2022). Some properties and applications of Half-Cauchy exponential extension distribution. *Int J Stat Appl Math;7*(6):91-101.

Telee, L. B. S., & Kumar, V. (2023). Modified Generalized Exponential Distribution. *Nepal Journal of Mathematical Sciences, 4*(1), 21–32. https://doi.org/10.3126/njmathsci.v4i1.53154

World Health Organization, WHO (2018). 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. *WHO. INT. 2018.* https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action.

World Health Organization. (2023). *WHO ambient air quality database, 2022 update: status report*. World