IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

A Statistical Analysis Of Air Pollution Using Modified Inverse Exponentiated Exponential Poisson Distribution Through Survival Analysis

[1]Monica.S, [2]Kavitha. S
[1]Ph.D Research Scholar, [2]Assistant Professor
[1]Department of Statistics
[1]Periyar University, Salem-11, India

Abstract : This study uses the Modified Inverse Exponentiated Exponential Poisson distribution, a special and adaptable probability distribution, to analyze the air quality in Kathmandu. A real-world dataset was used to evaluate the model's suitability for 2022–2027 air quality data in Kathmandu, Nepal. We evaluated current ground-level air quality conditions by analyzing data for particulate matter 2.5 (P2.5). To generate this distribution, an extra shape parameter is added to the Inverse Exponentiated Exponential Poisson distribution. Numerous statistical attributes of the proposed model are deduced and scrutinized. Furthermore, the parameter estimation, model validation, and model comparisons of the suggested model with current models for P2.5 at Ratnapark station, Kathmandu, have been investigated. Plotting the P-P and Q-Q charts verifies the model's graphical validation after the parameters are estimated using Maximum Likelihood estimation (MLE). The Anderson-Darling (An), Cramer-Von Mises (CVM), and Kolmogorov-Smirnov (KS) tests are used to compare the models' suitability and show that the suggested model fits data better.
The corresponding p-values are also provided. PM2.5 air pollution time series data from 2022 to 2027 are analyzed using the ARIMA model. Additionally, we discovered that the fundamental exponential smoothing approach offers a respectable forecast model for PM2.5 data spanning 2022–2027.

Index Terms: Hazard rate function, exponential distribution, and inversely exponentiated exponential Poisson distribution

I.Introduction

In recent years, fresh model classes have been developed to solve this issue by changing the standard Exponential Distribution. Lifetime distributions are crucial in fields like survival studies, actuarial science and reliability analysis, among others. The Exponential Distribution (ED) provides a simple and elegant analytical solution for many problems in lifetime testing and reliability studies. In the field of statistics and probability, theory creation of new lifetime distributions is necessary for modeling and analyzing lifetime data pertaining to the longevity or duration of various phenomena. However, the Exponential distribution falls short when it comes to fitting some real-world circumstances where hazard rates are not constant and display monotonic patterns. A generalized Exponential Distribution, sometimes called an Exponentiated Exponential Distribution was introduced by Gupta and Kundu (1999) [14] and expanded its use to data with failure rate functions that increase and decrease based on the shape parameter. Compounding the exponential distribution with different discrete distributions has been utilized by several reaches to create novel probability models. An exponential and geometric distribution were combined in the published exponential-geometric model by Adamidis and Loukas, 1998) [1], which showed a declining failure rate. The Exponential-Poisson (EP) distribution, which has two parameters and a decreasing failure rate, developed by Kus (2007) [20]. This distribution is the result of integrating a zero-truncated Poisson distribution with an exponential distribution. The Exponential-Poisson distribution's cumulative distribution function (or Cumulative Distribution Function) is

$$F_{kp}(x;\beta,\vartheta) = \frac{1}{(1-e^{-\vartheta})}[1-exp\{-\vartheta(1-e^{-\beta x})\}]\ ;\ x>0, \beta>0, \vartheta>0 \qquad (1.1)$$

developed by Barreto-Souza and Cribari- Neto (2009) [4] extended the distribution developed by (Kus,2007) [20] by adding a shape parameter. This extension called the Generalized Exponential Poisson distribution, has failure rates that exhibit decreasing, increasing, and upside-down bathtub patterns. The Cumulative Distribution Function of the Generalized Exponential Poisson distribution can be formulated as

$$FRP(x;\beta,\vartheta,\alpha) = [Fkp\alpha(x;\beta,\vartheta)]^{\alpha} \qquad (1.2)$$

$$= \frac{1}{(1-e^{-\vartheta})^{\alpha}}[1-exp\{-\vartheta(1-e^{-\beta x})\}]^{\alpha}\ ;\ x>0, \beta>0, \vartheta>0, \alpha>0 \qquad (1.3)$$

In a similar vein, Gupta and Kundu (2001) [15] developed the exponentiated exponential family while Cancho et al. (2011) [7] established a novel distribution family derived from the exponential distribution, distinguished by an increasing failure rate function. The Poisson Exponential (PE) distribution is capable of accurately representing lifetime data with increasing, decreasing, and upside-down bathtub-shaped failure rates. Bakouch et al. (2012) [4] introduced the Exponentiated Exponential Binomial distribution which combines features from exponentiated exponential and binomial distributions. The Exponentiated Weibull-Poisson (EWP) distribution is a unique distribution that was first introduced by Mahmoudi and Sepahdar (2013) [23]. Four distinct parameters describe this distribution which displays several failure rate patterns, including growing, decreasing, bathtub-shaped and unimodal. The Exponentiated Exponential Poisson distribution was created by Ristic and Nadarajah (2014) and has three different parameters. Depending on the shape parameter values, the Exponentiated Exponential Poisson distribution shows decreasing, rising, and upside-down bathtub-shaped failure rates. It is possible to formulate the unconditional Cumulative Distribution Function of the Exponentiated Exponential Poisson distribution as

$$Fkkp(x;\beta,\vartheta,\alpha)=1/(1-e^{-\vartheta})[1-exp\{-\vartheta(1-e^{-\beta x})\}]^{\alpha}\ ;\ x>0, \beta>0, \vartheta>0, \alpha>0 \qquad (1.4$$

Additionally, The Cumulative Distribution Function of the PNHE distribution can be formulated as follows: Chaudhary and Kumar (2020) [8] developed the Poisson NHE distribution with exhibiting rising, constant, and a diverse range of monotone failure rates; the Pareto Poisson–Lindley Distribution.

$$F_{pn}k(x;\beta,\vartheta,\alpha) = 1 - \frac{1}{(1-e^{-\vartheta})}[1-e^{-\beta x})^{\alpha}\ exp\{-\vartheta(1 - \}] \qquad (1.5)$$

$$= 1 - \frac{1-exp\,\{-\vartheta exp\,[1-(1+\alpha x)^{\alpha}]\}}{(1-e^{-\vartheta})} \; ; x>0, \beta > 0, \vartheta > 0, \alpha > 0 \qquad (1.6)$$

Similar life-time probability distributions have been created by other authors, including the Poisson Extended Exponential Distribution (Maya et al.), the Exponential Intervened Poisson and distribution (Jayakumar & Sankaran, 2021) [15], the Poisson Inverse NHE Distribution (Chaudhary & Kumar, 2020) [8], and the New Exponential Extension Poisson Distribution (Joshi and Kumar, 2021). In addition, Sousa-Ferreira and associates. A novel three-parameter lifetime distribution known as the Extended Chen-Poisson Lifetime Distribution was presented in 2023 [32]. In order to handle a range of intricate hazard shapes, this distribution combines the Chen distribution with the zero-truncated Poisson distribution.

A novel lifespan distribution called the Inverse Exponentiated Exponential Poisson distribution, which takes the inverse of the random variable X in EEP of (3), was proposed by Telee and Kumar (2023). The Inverse Exponentiated Exponential Poisson distribution's CDF may be written as

$$F_{IEEP}(x;\beta,\vartheta,\alpha) = 1 - \frac{1}{(1-e^{-\vartheta})} exp\{-\vartheta(1 - [1 - e^{-\beta/x})^{\alpha}\}] \; ; x>0, \beta > 0, \vartheta > 0, \alpha > 0 \qquad (1.7)$$

In addition to showing an upside-down bathtub-shaped hazard rate, the hazard rate function of the Inverse Exponentiated Exponential Poisson distribution exhibits a unimodal pattern with segments that are monotonically growing and decreasing. The idea behind creating a new distribution that can be used in a variety of fields, such as engineering, health, reliability analysis, survival analysis and demography. It is a promising choice for a variety of statistical and scientific research because of its adaptability and extensive application. Thus, this papers main goal is to analyze data on the PM2.5 air quality dataset from Kathmandu, Nepal, and create a unique distribution. This project seeks to comprehend Kathmandu, Nepal's present air quality condition. In many major locations around the world, including Nepal's capital city of Kathmandu, air pollution is a serious environmental problem. Kathmandu, which is tucked away in a valley and encircled by the magnificent Himalayan Mountains, has a special set of air quality issues. Residents health and well-being are adversely affected by the high levels of air pollutants caused by the citys geography, fast development, and numerous pollution sources. Both natural and man-made sources contribute to the air pollution in Kathmandu. The winter months bring more pollution to the city, mostly because of weather patterns that keep pollutants near the ground. Furthermore, the city's fast industrialization and urbanization have raised pollution emissions from industry, construction and automobiles.This issue is exacerbated by ineffective public transit networks and inadequate garbage management. This pollution has far-reaching effects on the environment, especially its unique biodiversity and cultural history, as well as the health of Kathmandu's citizens. The inversion effect is an atmospheric phenomena that is characteristic of Kathmandu, Nepal's winter and pre- monsoon seasons.

When there is no rainfall, this phenomena prevents contaminants from spreading. As a result, during these times, PM2.5, PM10, and Total Suspended Particles concentrations increase significantly. There are more forest fires around the country at this time, pollution levels rise noticeably from March to April. A research by Lamichhane et al. (2023) found that for 153 of the 309 days of air quality monitoring, PM2.5 levels were higher above the National Ambient Air Quality Standards. The environment and human health are impacted by particulate matter, which are small particles or droplets in the air that can be breathed into the lungs. These particles come from a variety of sources, such as natural sources like dust and pollen, industrial processes, and vehicle emissions. Because smaller particles can enter the respiratory system more deeply, particulate matter's size has a significant influence on health. Particulate matter is separated into different fractions based on size. Because of their potential to induce respiratory and cardiovascular problems, PM10 (particles with a diameter of 10 micrometers or less) and PM2.5 (particles with a diameter of 2.5 micrometers or less) are particularly concerning among these fractions (WHO,2023). According to Shrestha (2021) Nepal's annual average PM2.5 concentration decreased from 37.8 μg/m3 in 2019 to 35.8 μg/m3 in 2020. The average, however, surprisingly and considerably increased to 56.5 μg/m3 in 2021 (through May).

Widespread forest fires that erupted all throughout Nepal and peaked at the end of March were the main reason of this rise. PM2.5 values in the simulated 24-hour average vary from 30 μg/m3 to 65 μg/m3. The 24-hour mean of these simulated PM2.5 levels is higher than the WHO limit of 25 μg/m3. The air in Kathmandu has much higher PM2.5 levels, especially in the arid winter month of December. The urban air quality in Kathmandu is especially detrimental at this time, claim Tuladhar et al. (2021). The Air Quality Index has a numerical scale that ranges from 0 to 500. Air pollution worsens and the health hazards rise in tandem with the Air Quality Index score. For example, acceptable air quality is indicated by an Air Quality Index of 50 or below, whereas severely dangerous air quality is indicated by a rating of 300 or higher. According to the 2022 World Health Statistics report, the majority of people on the planet breathe air that is higher than World Health Organisation standards. The detrimental consequences of air pollution are responsible for about 3.5 million deaths annually from cardiovascular illnesses, accounting for more than 20% of all fatalities. According to the World Health Organisation, the number of people exposed to air pollution is rising, particularly in low- and middle- income nations like Nepal, which supports the idea that air pollution causes health issues in these regions. A novel method known as the New Extended Kumaraswamy Exponential Distribution was developed by Chaudhary et al. (2024) to analyze data on the air quality in Kathmandu, Nepal. In the city of Asansol, West Bengal, India, Banerjee et al. (2024) employed statistical modeling to examine the connection between air pollution and meteorological indices. The many elements of this study are arranged using the following structure. The Modified Inverse Exponentiated Exponential Poisson distribution to describe its mathematical properties in below. Numerical calculations are performed using the R programming language, which helps to create a methodical approach that leads readers through an understandable and well-structured framework. This strategy improves comprehension and admiration of the study's methodology and findings.

II. Prelimnaries

2.1. Definitions

Exponential Distribution is a popular continuous distribution that is frequently used to model the intervals between events. The exponential distribution is then defined mathematically and
is obtained as the mean and expected value. The parameter λ>0 indicates that a continuous random variable Xx has an exponential distribution. X is exponential, or λ. the function of probability density.

$$F_x(X) = \begin{cases} \lambda e^{-\lambda x} & x>0 \\ 0 & otherwise \end{cases}$$

Poisson Distribution represents the number of discrete events, whereas the exponentialdistribution represents the interval between those discrete events. The time between events follows an exponential distribution with a mean of 1/λ if the number of events follows a Poisson distribution with an average rate (λ) per unit of time. These two variables are both parts of the same Poisson process.

Cumulative Distribution Function is a random variable's likelihood of taking on a value less than or equal to a specified point .It applies to both discrete and continuous random variables and is a basic idea in probability and statistics. The CDF is the integral of the probability density function (PDF) for continuous variables from negative infinity to the point of interest.

$F_X(x) = P(X \leq x)$

A family of probability distributions known as the Generalized Exponential Distribution expands on the standard exponential distribution by including parameters to more accurately represent data from the real world.

The cumulative distribution function (CDF) in a two-parameter distribution is a popular form.

$F(x;\alpha,\lambda)=(1-e-\lambda x)\alpha \; ; \; \alpha, \lambda > 0$

in which $\alpha = 1$ and includes a standard exponential distribution. In contrast to the constant hazard rate of the conventional exponential distribution, this generalization permits a more flexible hazard rate that may be increasing, decreasing, or otherwise shaped.

Particulate matter is a complex mix of tiny solid particles and liquid droplets suspended in the air, including dust, soot, smoke, and fumes. It is categorized by its size, with the most harmful types being PM2.5 (particles 2.5 micrometers or smaller) and PM10 (particles 10 micrometers or smaller). These particles originate from various sources, like vehicle emissions and industrial processes, and can pose serious health risks by infiltrating the lungs and bloodstream.

2.1.2. Advantages of Exponentiated Exponential Distribution

- The inclusion of an extra parameter allows the EEP distribution to model more complex data sets and a wider range of failure rate shapes than simpler distributions like the exponential and Weibull distributions.
- The EEP distribution has been successfully applied to real-life data sets in engineering and other scientific fields, such as modeling software reliability and analyzing fatigue life and cancer remission time data.

- Several existing distributions are special cases of the EEP distribution. For example, the exponential-Poisson and exponential distributions are sub-models, which demonstrates its generalizing capacity

2.1.3. Disadvantages of Exponentiated Exponential Distribution

- Estimating the three parameters of the EEP distribution is more complex than for a simpler model. This is typically done using maximum likelihood estimation (MLE), which requires advanced statistical software like R and may require numerical methods to be solved.

- The underlying Poisson distribution assumes that its mean and variance are equal (equidispersion). If the data's variance is significantly greater than its mean (a common phenomenon known as overdispersion), using a model with a Poisson component can lead to inaccurate inferences.

- Like other complex, exponentiated distributions, the asymptotic convergence of the maximum likelihood estimators can be slow, especially with smaller sample sizes. This can make asymptotic inferences, such as confidence intervals, less accurate unless the sample size is very large.

2.1.4. Difference Between PM2.5 and P2.5 Air Pollution

	PM2.5	P2.5
Meaning	Particulate Matter with a diameter of 2.5 micrometers or less.	This is not a standard air quality term. It is a common typo or misinterpretation of PM2.5.
Relevance to air quality	A standard, medically-relevant measure of air pollution recognized by health and environmental agencies globally.	Not relevant to air quality.
Health implications	Because of its small size, PM2.5 can penetrate deep into the lungs and enter the bloodstream, leading to serious health issues like heart and lung disease.	N/A
Use in other industries	N/A	"P2.5" correctly refers to the pixel pitch of an LED display screen. The "P" signifies "pixel," and 2.5mm is the distance between pixels.
Context	Health and environmental science.	Digital display and manufacturing.

3. Methods of Modified Inverse Exponentiated Exponential Poisson Distribution Formulation

To make the probability model more flexible, we have added a shape parameter, θ, to 5 of the Cumulative Distribution Function of the Inverse Exponentiated Exponential Poisson distribution as described by Telee & Kumar in 2023, this improved probability model is called the Modified Inverse Exponentiated Exponential Poisson distribution with three parameters (α, β, λ, θ) > 0, and its CDF is therefore expressed by

$$F(x;\alpha,\beta,\vartheta,\lambda) = 1 - \frac{1}{(1-e^{-\lambda})}\left[1 - \exp\left\{-\lambda\left(1 - e^{-\beta x^{-1}} e^{-\vartheta x}\right)^{\alpha}\right\}\right];\ x>0,\ \alpha>0,\ \beta>0,\ \vartheta>0 \qquad (3.1)$$

The associated PDF of equation is included as

$$f(x;\alpha,\beta,\vartheta,\lambda) = \frac{\alpha\beta\lambda}{(1-e^{-\lambda})} x^{-2} e^{-\vartheta x} v[1+\vartheta x](1-v)^{\alpha-1}\left(\exp\{-\lambda(1-v)^{\alpha}\}\right);\ \text{where, } v = e^{-\beta x^{-1}} e^{-\vartheta x} \qquad (3.1.2)$$

3.1. Survival function (S(x))

The survival function The survival function is used in fields like medicine and engineering to analyze events like death, disease occurrence, or equipment failure over time. The S(x) for the proposed Modified Inverse Exponentiated Exponential Poisson distribution model is represented by

$$S(x) = \frac{1}{(1-e^{-\lambda})}\left[1 - \exp\left\{-\lambda\left(1 - e^{-\beta x^{-1}} e^{-\vartheta x}\right)^{\alpha}\right\}\right];\ x>0,\ \alpha>0,\ \beta>0,\ \vartheta>0 \qquad (3.3)$$

Figure 1 & 2 shows probability density curves for α = 0.1657 and β = 42.9437 on the left, and pdf curves for β = 42.9437 and λ= 3 on the right. The Probability Density Function plots show a unimodal distribution with a positive skew, indicating that values are concentrated around a central point.

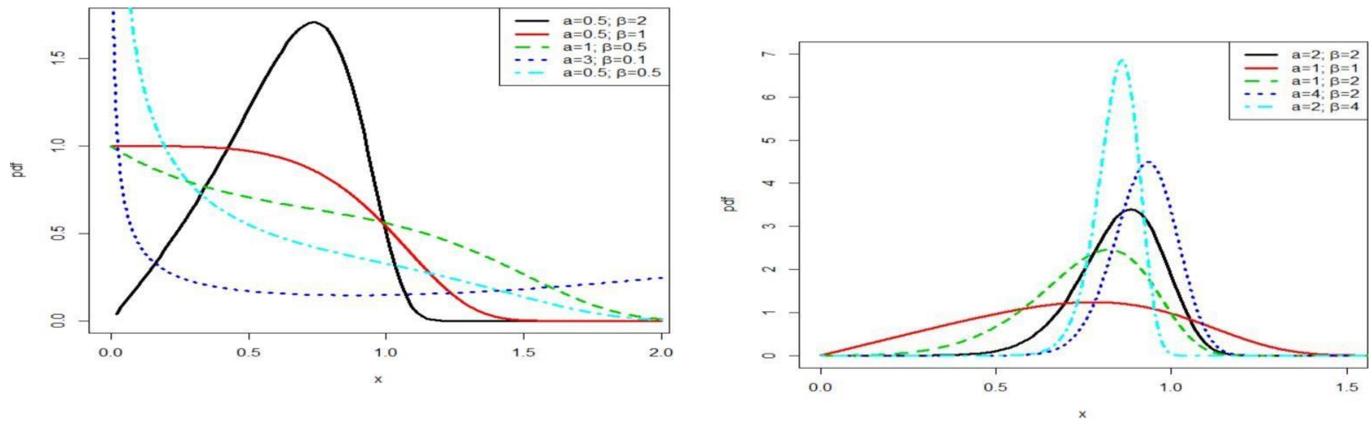


Fig1:Probability density function

3.2Hazard Rate Function (h(x))

The equation defines the hazard rate, which is the rate at which breakdowns occur instantaneously.

h(x) = $\alpha\beta\lambda[1+\vartheta x]x^{-2}e^{-\vartheta x}\ v(1-v)^{\alpha-1}[exp\{-\lambda(1-v)^{\alpha}\}][1-exp\{-\lambda(1-u)^{\alpha}\}]^{-1}$ (3.4)

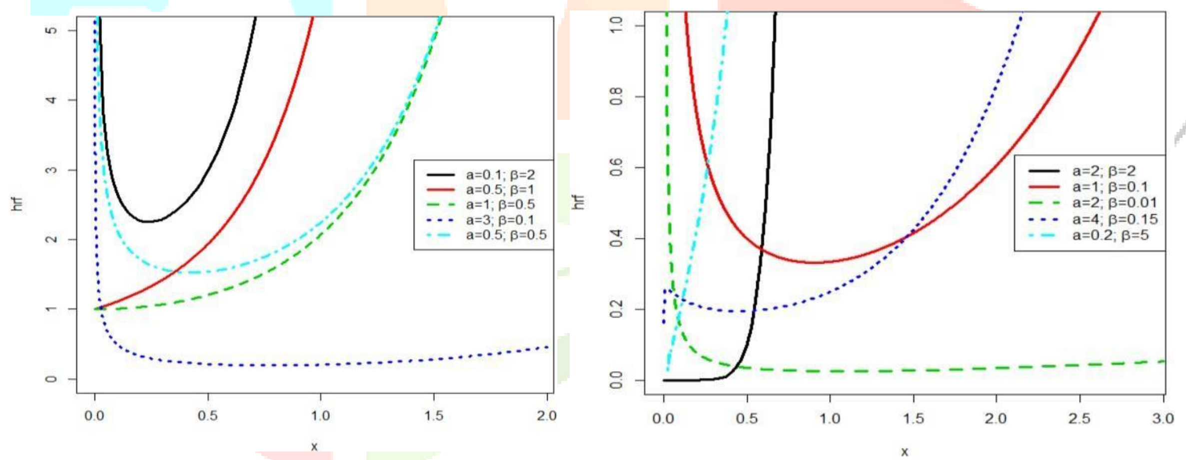


Figure 3 &4 illustrates the suggested model's hazard rate curves. The figure 3 of the hazard rate curve assumes β = 42 and ϑ = 0.1, whereas the right side assumes β = 42 and ϑ = 3. The plot of the hazard rate function displays both a rising trend and a unique inverted bathtub shape, showing changing risk levels over time.

3.3 Reversed hazard rate function

The reversed hazard rate, hrev(x), is defined by

−1

$\alpha\beta\gamma(1+\theta)x^{-2}e^{-\vartheta x}\ v(1-v)^{\alpha-1}[exp\{-\lambda(1-e^{v})^{\alpha}\}]\ [1-$ 1 $[1-exp\{-\lambda(1-v)^{\alpha}\}]]$ (3.5)

Hrev(x) =

$(1-e^{-\lambda})$ $(1-e^{-\lambda})$

3.4 Cumulative Hazard Rate

The expression (11) represents the cumulative hazard rate.

H(x)= ln[$\frac{1}{(1-e^{-\lambda})}$ $[1-exp\{-\lambda(1-e^{-\beta})^{\alpha}\}]]$; x>0, $\alpha>0,\beta>0,\vartheta>0,\lambda>0$

3.5 Quantile function

The quantile function for MIEEP is defined as

$$\frac{1}{\beta}e^{-\vartheta x} + \left(\frac{1}{\beta}\right)\log\left[1-\left\{\left(\frac{-1}{\lambda}\right)\log\left\{1-\left(1-p\left(1-e^{-\lambda}\right)\right)^{\frac{1}{\alpha}}\right\}\right\}\right] = 0\ ;\ 0\le p\le 1 \qquad (3.6)$$

3.6 Asymptotic Behavior

To evaluate the asymptotic behavior of the density function, use the following equation:

$\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x)$. If the model meets the asymptotic criteria, the modal value will be clearly indicated. Limits at the endpoints should also be considered.

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \frac{\alpha\beta\lambda x^{-2} u e^{-\vartheta x}}{(1-e^{-\lambda})}\left[1 + \vartheta x\right](1-v)^{\alpha-1}\left[exp\{-\lambda(1-v)^{\alpha}\}\right] = 0 \qquad (3.7)$$

$$\lim_{x\to \infty} f(x) = \lim_{x\to \infty} \frac{\alpha\beta\lambda x^{-2} u e^{-\vartheta x}}{(1-e^{-\lambda})}\left[1 + \vartheta x\right](1-v)^{\alpha-1}\left[exp\{-\lambda(1-v)^{\alpha}\}\right] = 0 \qquad (3.8)$$

In this Case, $\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x)$. implying that the mode of the model will exist.

3.7 Skewness

Al-Saiary et al. (2019) [2] gave quartile-based of skewness as

$$SKB = \frac{Q\left(\frac{3}{4}\right) - \frac{1}{2}\times Q\left(\frac{1}{2}\right) + Q\left(\frac{1}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}$$

3.8 Kurtosis

The Octiles coefficient of Kurtosis, as defined by (Moors, 1988), can be computed using the following formula:

$$K = \frac{Q\left(\frac{3}{4}\right) - Q(0.125) + Q(0.875) - Q(0625)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)} \qquad (3.9)$$

3.9 Methods for Estimation of Model Constants

The method of maximum likelihood estimation is used to estimate the parameters. This approach of estimating is based on maximizing the model's log likelihood. Assume we have a random sample of n observations from Modified Inverse Exponentiated Exponential Poisson distribution, denoted by $\bar{X} = (x_1, x_2, \ldots, x_n)$. The log likelihood function is expressed as follows:

$$f(x;\alpha,\beta,\vartheta,\lambda|\underline{x}) = n\log(\alpha\beta\lambda) - n\log(1-e^{-\lambda}) - \lambda\sum_{i=1}^{n}\left(1-exp\left(\frac{-\beta e^{-\vartheta x_i}}{\vartheta x_i}\right)\right)^{\alpha} - \beta\sum_{i=1}^{n}\frac{e^{-\vartheta x_i}}{x_i})$$

$$-\sum_{i=1}^{n}(\vartheta x_i - \log x^{-2}) + (\alpha-1)\sum_{i=1}^{n}\log\left(1-\exp\left(\frac{-\beta e^{-\frac{\vartheta}{x_i}}}{x_i}\right)\right) \qquad (3.10)$$

To determine the first and second order partial derivatives of the log-likelihood function, differentiate equation with respect to α, β, λ, and θ. We may estimate the suggested model's parameters by setting the first-order derivatives to zero and solving for them. However, if getting a solution for the first-order partial derivatives is not possible, computer programming can be used to solve the system of nonlinear equations.

III. Applications

To examine the model's flexibility and used the 2021 Real-time Air Quality Management System (AQMS) dataset. The dataset comprises particulate matter measured at minute intervals as provided by GoN/MoFE (2023). The following links were utilized to gather real-time air quality datasets for 2021 in Kathmandu, Nepal has given reference. The Nepalese Department of Environment constructed 27 Air Quality Monitoring Stations around the nation in 2021 to use PM2.5 data to determine the actual state of the country's air quality. This article analyzes 2021 PM2.5 data from the Ratnapark Station in Kathmandu, Nepal. Additionally, we performed time series analysis on P2.5 air quality data from the Ratnpark station in Kathmandu of P2.5 data were utilized to assess and show summary statistics, parameter estimations, model validations, and model comparisons using R software's optim () function PM2.5 is defined by the WHO as particulate matter with an aerodynamic diameter of 2.5 µm or less. Health risks may arise if air quality data falls outside of these ranges.

1.2 Data Analysis for P2.5 Air Quality Standards

The Nepalese government's legal air quality requirements are the National Ambient Air Quality requirements, 2012 (NAAQS) certain parameters' upper limitations are specified in these recommendations (table 1).

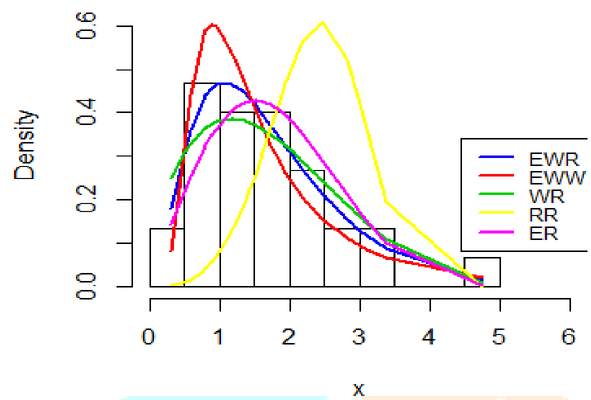
S.No	Parameters	Unit	Averaging Time	Minimum Concentration
1	Lead	$\mu g/m^3$	Annual	0.5
2	Carbon Monoxide	$\mu g/m^3$	8 hour	10000
3	Bonzene	$\mu g/m^3$	Annual	5
4	NO2	$\mu g/m^3$	Annual	50
			24 hour	70
5	SO2	$\mu g/m^3$	Annual	60
			24 hour	80
6	TSP	$\mu g/m^3$	24 hour	240
7	Ozon	$\mu g/m^3$	8 hour	163
8	PM10	$\mu g/m^3$	24 hour	110
9	PM21	$\mu g/m^3$	24 hour	50

Table 2: Summary of the data reveals that the mean P2.5 value at the Ratnapark Station is higher than the daily median value

Mean	Q1	Q2	Q3	Maximum	Minimum	Skewness	SD	Kurtosis
49.978	17.200	37.321	68.271	207.021	6.988	1.721	34.975	5.767

Table 3: The maximum likelihood estimation model was used to estimate the parameters' values.

Parameters	MLE	SE
A	0.1467751	0.0315277
B	41.432786	4.0673912
Λ	0.3462819	0.2917635
ϑ	0.1533621	0.0308451



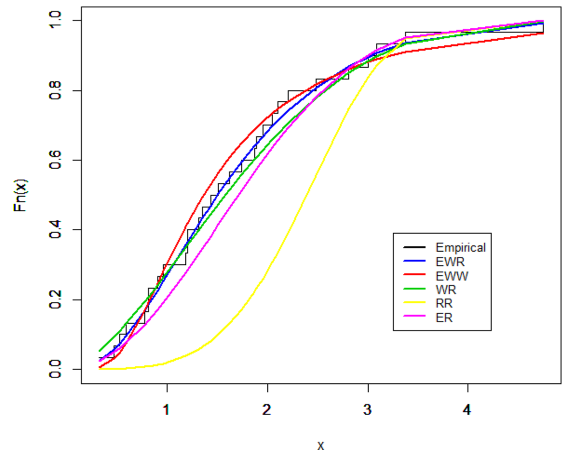
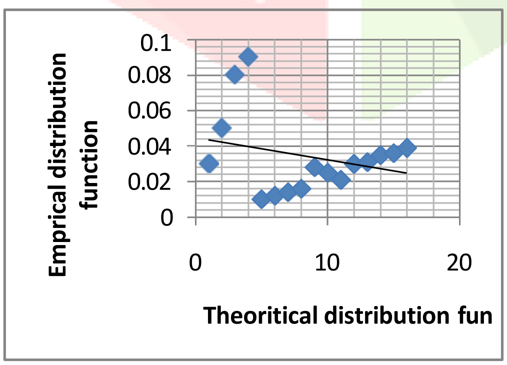


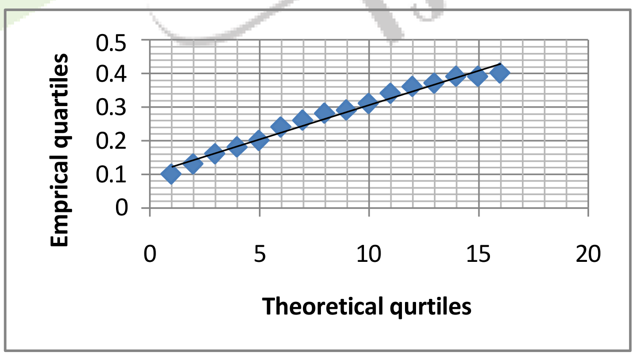
Table 4 shows the associated p-values for the different estimate methodologies together with the goodness of fit test statistics values for Anderson-Darling (A2), Cramer-von Mises (W), and Kolmogorov-Smirnov (KS).

Table4: A2, W, and KS statistics with associated p- values

Methods	Critical values	p-value
Kolmogorov-Smirnov	0.0643286	0.076081
Cramer-von Mises	0.2917561	0.100739
Anderson-Darling	2.5743890	0.041723

We also created P-P and Q-Q plots for the suggested model to assess its validity, which are shown in figure 6.





Modified Inverse Exponentiated Exponential Poisson (MIEEP) distribution, Generalized Exponential Extension (GEE) distribution, Half Logistic Nadarajah Haghighi (HLNHE) distribution, the Lindley Inverse Weibull (LIW) distribution, Weibull Extension (WE) distribution, Generalized Weibull Extension (GWE) distribution

The parameters of the models in question, which were calculated using MLE, are shown in Table 5. The SE for each of the examined distributions is also displayed in Table 5.

Table 5: SE and estimated parameters for NEKwE, along with an assessment

Model	α	B	θ	Λ
MIEEP	0.175	4.089	0.598	0.4876
GEE	0.763	4.632	-	0.206
LIW	19.885	0.139	14.198	-
HLNHE	0.819	0.068	-	0.009
WE	0.005	0.004	-	0.185
GWE	52.168	0.0187	-	52.749

The models under evaluation and the information criterion values for Modified Inverse Exponentiated Exponential Poisson are displayed in Table 6. The data appears to match the proposed model better than the tested models as the information criteria values for the proposed model are lower than those for the analyzed model.

Model	-LL	AIC	BIC	CAIC	HQIC
MIEEP	1423.910	2819.775	2827.863	2806.520	2819.147
GEE	1617.301	3375.605	3378.150	3358.050	3363.919
LIW	1618.108	3375.809	3386.143	3361.129	3374.011
HLNHE	1618.461	3375.920	3396.105	3381.534	3389.677
WE	1634.873	3391.465	3399.725	3386.099	3393.853
GWE	1647.664	3408.180	3463.541	3424.104	3451.062

-LL – Negative Log Likelihood AIC - Akaike Information Criterion
.BIC - Bayesian Information Criterion
CAIC - Corrected Akaike Information Criterion HQIC - Hannan-Quinn Information Criterion

1.3 Holt-Winters Exponential Smoothing

Holt-Winters exponential smoothing works well for short-term forecasting exhibits an additive pattern with a trend (either upward or downward) and seasonal variability. This approach determines the time series' current level, trend rate, and seasonal effect. Alpha, beta, and gamma are the three main parameters involved. Gamma corrects for the seasonal component's fluctuation during the present period, beta regulates the smoothing of the trend slope, and alpha establishes the smoothing factor for the level. The values of alpha, gamma, and beta range from 0 to 1. Recent observations are given very little weight in future value estimates when these values are near zero.
Addition of seasonal components and trend in Holt-Winters exponential smoothing Parameters for smoothing:

α is calculated to be 0.0381, which is a very low value. This implies that recent measurements as well as certain observations from earlier times are crucial to the present level estimate. When beta stays at 0.00, it means that the trend component's slope estimate stays at its starting value for the duration of the time series at 0.498, γ is noticeably high, indicating that the most recent observations have a significant impact on the estimation of the seasonal component at this time point. The graphic shows that the seasonal peaks, which usually appear in the fourth quarter of the year, are well predicted by the Holt-Winters exponential approach.
According to the research, the expected errors are probably normally distributed, meaning that their mean is zero. Additionally, there is little indication of non-zero autocorrelation in the forecast errors during the sample period, according to the Ljung-Box test. This supports the finding that, with a mean of zero, the forecast error distribution resembles normalcy. These results suggest that the basic exponential smoothing approach provides a good forecast model for the 2022–2026.

Year/Qutr	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2024 Q1	68.173724	41.0229154	94.07863	25.680541032	107.90451
2024 Q2	32.534749	6.018600	58.98765	-8.254109842	73.321550
2024Q3	11.322516	-17.091042	37.32568	-31.03487150	51.953118
2024Q4	63.908663	35.9111304	90.12564	22.098361555	105.32654
2025Q1	64.432439	35.5487632	95.21541	18.203452100	111.39013
2025Q2	27.617432	-0.703281	60.32677	-16.396000101	76.7750
2025 Q3	7.135486	-23.715086	37.88053	-39.875426672	54.01446
2025 Q4	61.934524	31.0632458	91.48010	14.2067785431	107.13300
2026 Q1	62.312075	28.0078143	94.76105	11.6345901163	113.00418
2026 Q2	27.410993	-6.1765531	61.43210	-24.061883250	78.83457
2026 Q3	5.709021	-27.908637	39.21000	-46.483306518	55.88766
2026 Q4	57.108899	25.2176581	92.05431	6.478903211	110.41018
2027 Q1	59.265421	23.0963214	96.40507	3.976320180	115.23139
2027 Q2	23.907086	-12.864311	61.30548	-31.80155459	80.20918
2027 Q3	2.51632	-32.917632	39.32564	-53.284110861	58.092155
2027 Q4	56.495782	18.099176	92.24867	-0.0091054328	111.03311

IV. Conclusion

The Modified Inverse Exponentiated Exponential Poisson distribution is a new and flexible probability distribution that has been presented in this work and investigated this novel model's many statistical characteristics in detail. Examine a real dataset of P2.5 air quality readings from the Ratnapark station in Kathmandu, Nepal, throughout the course of 2027 in order to assess the suitability of our model.To estimate parameters, verify models, and compare models for P2.5 data and conducted data analysis.
However, P2.5 concentrations showed seasonal changes, peaking during the winter and falling during the monsoon season. To validate our model, we employed a number of diagnostic techniques, such as P-P and Q-Q graphs. Furthermore, we assessed metrics like the Bayesian information criterion, the Hannan- Quinn information criterion, the corrected Akaike information criterion, log-likelihood values, and the Akaike information criterion conducted Anderson-Darling, Cramer-Von Mises, and Kolmogorov- Smirnov tests to evaluate the model's accuracy and appropriateness. Additionally, we performed a time series analysis of P2.5 air pollution data from 2024 to 2027 at Kathmandu, Nepal's Ratnapark. Our research showed that, aside from the COVID-19 epidemic, the air quality in Kathmandu, Nepal, was continuously bad. Furthermore, we found that PM2.5 levels from 2022 to 2027 may be accurately predicted using the straightforward exponential smoothing approach.

V. Reference

1. Chaudhary AK, Telee LBS, Karki M, Kumar V. Statistical analysis of air quality dataset of Kathmandu, Nepal, with a New Extended Kumaraswamy Exponential Distribution. Environmental Science and Pollution Research; c2024 .p. 1-16. https://doi.org/10.1007/s11356-024-32129-z

2. Lamichhane GP, Maharjan N, Pandey B, K.C. P, Paudel R, Khanal S. Status of Air Quality in Nepal, Annual Report, 2021. Department of Environment, Kathmandu, Nepal; c2023

3. Asgharzadeh A, Bakouch HS, Esmaeili L. Pareto Poisson–Lindley distribution with applications. Journal of Applied Statistics. 2013;40(8):1717-1734.

4. Banerjee B, Kundu S, Kanchan R, Mohanta A. Examining the relationship between atmospheric pollutants and meteorological factors in Asansol city, West Bengal, India, using statistical modelling. Environmental Science and Pollution Research; c2024 .p. 1-18. https://doi.org/10.1007/s11356-024-33608-z

5. Cancho VG, Louzada-Neto F, Barriga GD. The Poisson-exponential lifetime distribution. Computational Statistics and Data Analysis. 2011;55(1):677-686. https://doi.org/10.1016/j.csda.2010.05.033

6. Chaudhary AK, Kumar V. Poisson NHE Distribution: Properties and Applications. International Journal of Applied Research. 2020;6(12):399-409. https://doi.org/10.22271/allresearch.2020.v6.i12f.8143

7. Chaudhary AK, Kumar V. Poisson Inverse NHE Distribution with Theory and Applications. International Journal of Science and Research. 2020;9(12):1603-1610. DOI: 10.21275/SR201226151601

8. Fatima A, Roohi A. Extended Poisson exponential distribution. Pakistan Journal of Statistics and Operation Research. 2015:361-375. https://doi.org/10.18187/pjsor.v11i3.708

9. GoN/MoFE. Air Quality Management System (AQMS) Data-2021. Department of Environment, Babarmahal, Kathmandu, Nepal; c2023. http://doenv.gov.np/progressfiles/Final-data-compilation_2021_17-stations- 1690193536.xlsx

10. Lamichhane GP, Maharjan N, Pandey B, K.C. P, Paudel R, Khanal S. Status of Air Quality in Nepal, Annual Report, 2021. Department of Environment, Kathmandu, Nepal; c2023.

11. Lemonte AJ. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. Computational Statistics and Data Analysis. 2013;62:149-170. https://doi.org/10.1016/j.csda.2013.01.011

12. Louzada F, Marchi V, Roman M. The exponentiated exponential–geometric distribution: a distribution with decreasing, increasing and unimodal failure rate. Statistics. 2014;48(1):167-181. https://doi.org/10.1080/02331888.2012.667103

13. Mansoor M, Tahir MH, Alzaatreh A, Cordeiro GM. The Poisson Nadarajah–Haghighi distribution: Properties and applications to lifetime data. International Journal of Reliability, Quality and Safety Engineering. 2020;27(01):2050005. https://doi.org/10.1142/S0218539320500059

14. Muhammad M. Poisson-odd generalized exponential family of distributions: Theory and applications. Hacettepe Journal of Mathematics and Statistics. 2018;47(6):1652-1670.

15. Shrestha SL. Analysis of ambient particulate air pollution and health in NEPAL. Journal of Global Ecology and Environment. 2021;12(4):14-29. https://doi.org/10.29121/ijoest.v7.i5.2023.484

16. Waqas H, Yousaf Z, Shafiq Z, Kareem A, Nadeem S. On a new generalization of exponential–Poisson distribution with application to wind speed data. Mathematics. 2022;10(12):2055.

Dataset Link 1: https://doi.org/10.3390/math101220551690193536.xlsx Dataset Link 2:

http://doenv.gov.np/progress/6/28416357