

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(4): 125-138
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 28-06-2024
Accepted: 01-08-2024

Arun Kumar Chaudhary
Department of Statistics, Nepal
Commerce Campus, Tribhuvan
University, Kathmandu, Nepal

Lal Babu Sah Telee
Department of Statistics, Nepal
Commerce Campus, Tribhuvan
University, Kathmandu, Nepal

Murari Karki
Department of Statistics,
Sarswati Multiple college,
Tribhuvan University,
Kathmandu, Nepal

Vijay Kumar
Department of Mathematics and
Statistics, DDU Gorakhpur
University, Gorakhpur, Uttar
Pradesh, India

Corresponding Author:
Arun Kumar Chaudhary
Department of Statistics, Nepal
Commerce Campus, Tribhuvan
University, Kathmandu, Nepal

Modified inverse exponentiated exponential Poisson distribution to Analyze air quality dataset of Kathmandu, Nepal

Arun Kumar Chaudhary, Lal Babu Sah Telee, Murari Karki and Vijay Kumar

DOI: <https://dx.doi.org/10.22271/math.2024.v9.i4b.1783>

Abstract

This article analyzes the air quality status of the Kathmandu city using a novel and flexible probability distribution called the Modified Inverse Exponentiated Exponential Poisson distribution. This distribution is crafted by incorporating an extra shape parameter into the Inverse Exponentiated Exponential Poisson distribution. Several statistical features of the suggested model are derived and analyzed. A real data set was utilized to assess the model's suitability for the air quality data of Kathmandu, Nepal, spanning the years 2017 to 2021. In order to examine the actual ground-level air quality conditions, we conducted the data analysis for particulate matter 2.5 ($P_{2.5}$). Furthermore, we have studied the parameter estimation, model validation and model comparisons of proposed model with existing models for $P_{2.5}$ at Ratnapark station, Kathmandu. The model's parameters are estimated through Maximum Likelihood estimation (MLE). The P-P and Q-Q charts are plotted to check the graphical validation of the model. To compare the suitability of the models, Akaike, Corrected Akaike, Hannan-Quinn and Bayesian Information criteria are used. Anderson-Darling (An) test, Cramer-Von Mises (CVM) test, Kolmogorov-Smirnov (KS) test, along with their corresponding p-values are used to show the better fitting of the proposed model. ARIMA model is applied for time series analysis of air pollution data of $PM_{2.5}$ from during years 2017-2021. We also discovered that the basic exponential smoothing approach gives a good forecast model for $PM_{2.5}$ data from 2017 to 2021.

Keywords: Exponential distribution, inverse exponentiated exponential Poisson distribution, hazard rate function, holt-winters exponential smoothing, $p_{2.5}$ air quality

1. Introduction

In the field of statistics and probability, the development of new lifetime distributions is essential for modelling and analyzing lifetime data related to the longevity or duration of various phenomena. Lifetime distributions play a critical role in areas such as reliability analysis, survival studies, and actuarial science, among others. The exponential distribution (ED) offers a straightforward and elegant analytical solution for numerous challenges encountered in lifetime testing and reliability studies. However, the ED falls short when it comes to fitting certain real-world scenarios where hazard rates are not constant and exhibit monotonic patterns. In recent years, novel model classes have been developed to address this issue by modifying the traditional ED.

Gupta and Kundu (1999) ^[14] presented a generalized ED (also known as exponentiated exponential distribution) that broadens its application to data with growing and decreasing failure rate functions dependent on the shape parameter. Many authors to generate new probability models have used compounding of the exponential distribution with various discrete distributions. Exponential-geometric model (Adamidis and Loukas, 1998) ^[1] was published combining an exponential & geometric distribution, demonstrating a decreasing failure rate. In a similar vein, Kus (2007) ^[20] introduced the exponential-Poisson (EP) distribution, defined by two parameters and exhibiting a decreasing failure rate. This distribution arises from the integration of an exponential distribution with a zero-truncated Poisson distribution.

The cumulative distribution function (CDF) for EP distribution can be expressed as

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

Barreto-Souza and Cribari-Neto (2009) ^[4] improved the distribution provided by (Kus,2007) ^[20] by include a shape parameter. This modification increased the distribution's versatility in modeling real-world data. This extension, known as the Generalized Exponential Poisson (GEP) distribution, features failure rates that demonstrate decreasing, increasing, and upside-down bathtub patterns. The CDF of GEP distribution can be formulated by (2).

$$F_{GEP}(x; \beta, \lambda, \alpha) = [F_{EP}(x; \beta, \lambda)]^\alpha = \frac{1}{(1 - e^{-\lambda})^\alpha} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta x} \right) \right\} \right]^\alpha; x > 0, \lambda > 0, \beta > 0, \alpha > 0 \quad (2)$$

In a similar fashion, Gupta & Kundu (2001) ^[15] created Exponentiated exponential family and Cancho *et al.* (2011) ^[7] introduced a new distribution family derived from the exponential distribution, which is characterized by its increasing failure rate function. This distribution is known as the Poisson exponential (PE) distribution. Bakouch *et al.* (2012) ^[4] introduced the Exponentiated Exponential Binomial (EEB) distribution, blending characteristics from exponentiated exponential and binomial distributions. This distribution is capable of accurately representing lifetime data characterized by increasing, decreasing, and upside-down bathtub-shaped failure rates.

Mahmoudi and Sepahdar (2013) ^[23] introduced a novel distribution known as the Exponentiated Weibull-Poisson (EWP) distribution. This distribution is defined by four specific parameters and shows different patterns of failure rates: increasing, decreasing, bathtub-shaped, and unimodal.

Ristic & Nadarajah (2014) ^[29] developed Exponentiated exponential Poisson (EEP) distribution characterized by three distinct parameters, exhibiting decreasing, increasing and upside-down bathtub shaped failure rates depending upon the shape parameter values. The unconditional CDF of EEP distribution can be formulated as (3)

$$F_{EEP}(x; \beta, \lambda, \alpha) = \frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta x} \right)^\alpha \right\} \right]; x > 0, \lambda > 0, \beta > 0, \alpha > 0 \quad (3)$$

Furthermore, the known lifetime compounded distributions created with exponential distribution and various discrete probability distributions by various writers are Pareto Poisson–Lindley Distribution (Asgharzadeh *et al.*,2013) ^[3], Exponentiated Exponential–Geometric Distribution (Louzada *et al.*,2014) ^[22], Extended Poisson exponential distribution (Fatima & Roohi, 2015) ^[11], Poisson-odd generalized exponential family of distributions (Muhammad,2018) ^[27], Poisson Nadarajah–Haghighi Distribution (Mansoor *et al.*,2020). Likewise, Chaudhary and Kumar (2020) ^[8] developed Poisson NHE distribution with exhibiting rising, constant and a diverse range of monotone failure rates. The CDF of PNHE distribution can be formulated as

$$F_{PNHE}(x; \beta, \lambda, \alpha) = 1 - \frac{1 - \exp \{ -\lambda \exp [1 - (1 + \alpha x)^\beta] \}}{(1 - e^{-\lambda})}; x > 0, \lambda > 0, \beta > 0, \alpha > 0 \quad (4)$$

In the same way, life time probability distributions developed by various authors are New Exponential Extension Poisson Distribution (Joshi & Kumar, 2021) ^[18], Poisson Inverse NHE Distribution (Chaudhary & Kumar,2020) ^[8], Exponential Intervened Poisson and distribution (Jayakumar & Sankaran, 2021) ^[15], Poisson Extended Exponential Distribution (Maya *et al.*,2022). Sousa-Ferreira *et al.* (2023) ^[32] introduced a new three-parameter lifetime distribution, named the Extended Chen-Poisson Lifetime Distribution. This distribution is created by combining the Chen distribution with the zero-truncated Poisson distribution, allowing it to accommodate various complex hazard shapes.

Telee and Kumar (2023) ^[34] recommended new lifetime distribution known as Inverse Exponentiated Exponential Poisson (IEEP) distribution taking inverse of random variable X in EEP of (3). The CDF of IEEP distribution can be expressed as

$$F_{IEEP}(x; \beta, \lambda, \alpha) = 1 - \frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta/x} \right)^\alpha \right\} \right]; x > 0, \lambda > 0, \beta > 0, \alpha > 0 \quad (5)$$

The hazard rate function (HRF curve) of IEEP distribution demonstrates a unimodal pattern, characterized by both monotonically increasing and monotonically decreasing segments as well as it displays an upside-down bathtub-shaped hazard rate.

The concept of developing a new distribution entail introducing a flexible distribution that can be applied in diverse areas including health, engineering, demography, survival analysis, and reliability analysis. Its flexibility and broad applicability make it a promising option for a wide range of statistical and scientific investigations. Therefore, the primary objective of this paper is

to develop a novel lifetime distribution and perform data analysis on the PM_{2.5} air quality dataset from Kathmandu, Nepal. This innovative initiative aims to understand the current air quality situation in Kathmandu, Nepal.

Air pollution is a significant environmental concern in many urban areas worldwide, and Kathmandu, the capital city of Nepal, is no exception. Nestled in a valley surrounded by the majestic Himalayan Mountains, Kathmandu faces a unique set of challenges when it comes to air quality. The city's topography, rapid urbanization, and a variety of pollution sources contribute to elevated levels of air pollutants, negatively impacting the health and well-being of its residents. Kathmandu's air pollution results from a combination of natural and anthropogenic factors. The city experiences increased pollution during the winter months, primarily due to meteorological conditions, which trap pollutants close to the ground. Moreover, the rapid urbanization and industrialization of the city have led to increased emissions of pollutants from vehicles, industries, and construction activities. Inefficient public transportation systems and a lack of proper waste management contribute to this problem. The consequences of this pollution are far-reaching, affecting not only the health of Kathmandu's residents but also the environment, including its rich biodiversity and cultural heritage.

In Kathmandu, Nepal, the winter and pre-monsoon seasons are characterized by an atmospheric phenomenon called the inversion effect. This phenomenon impedes the dispersal of pollutants in the absence of rainfall. Consequently, concentrations of PM_{2.5}, PM₁₀, and TSP (Total Suspended Particles) rise dramatically during these periods. In addition, pollution levels experience a notable increase from March to April owing to heightened occurrences of forest fires nationwide during this period. According to a study conducted by (Lamichhane *et al.*, 2023)^[20], out of 309 days of air quality monitoring, PM_{2.5} levels exceeded the National Ambient Air Quality Standards on 153 days.

Particulate matter (PM) are microscopic particles or droplets in the air that can be inhaled into the lungs and affect human health and the environment. These particles originate from a range of sources, including industrial operations, automobile emissions, and natural sources like dust and pollen. The size of particulate matter is a critical factor in its impact on health, as smaller particles can penetrate deeper into the respiratory system. PM, or particulate matter, is divided into various fractions according to size. Among these fractions, PM₁₀ (particles with a diameter of 10 micrometres or less) and PM_{2.5} (particles with a diameter of 2.5 micrometres or less) are particularly worrisome due to their ability to potentially cause respiratory and cardiovascular issues, as documented by (WHO, 2023).

Shrestha (2021)^[31] reported that the yearly average of PM_{2.5} in Nepal dropped from 37.8 µg/m³ in 2019 to 35.8 µg/m³ in 2020. However, in 2021 (unexpectedly and significantly), the average rose to 56.5 µg/m³ (through May). This increase was primarily caused by widespread forest fires that occurred throughout Nepal, peaking at the end of March.

The simulated 24-hour average PM_{2.5} concentrations range from 30 µg/m³ to 65 µg/m³. These simulated PM_{2.5} values exceed the World Health Organization (WHO) standard of 25 µg/m³ for a 24-hour mean. The PM_{2.5} levels in the air of Kathmandu are notably elevated, particularly during the dry winter month of December. According to Tuladhar *et al.* (2021), the urban air quality in Kathmandu is particularly harmful during this time period.

A numerical scale of the Air Quality Index (AQI) lies between 0 and 500. As the AQI number goes up, air pollution gets worse, and the health risks also increase. For instance, an AQI of 50 or below means the air quality is excellent, whereas a reading over 300 indicates extremely hazardous.

World Health Statistics report of 2022 says that most people worldwide breathe air that surpasses WHO guidelines for air quality. More than 20% of deaths from cardiovascular diseases are caused by the harmful effects of air pollution, totalling over 3.5 million deaths each year. The WHO reports an increasing number of people exposed to air pollution, especially in low and middle-income countries like Nepal, which strengthens the evidence linking air pollution to health problems in these areas. Chaudhary *et al.* (2024)^[10] created innovative modal called a New Extended Kumaraswamy Exponential Distribution for conducting data analysis of air quality conditions of Kathmandu, Nepal. Banerjee *et al.* (2024) used statistical modelling to investigate the relationship between air pollution and climatic parameters in the city of Asansol, West Bengal, India.

The following framework is employed to organize the different components of this study. In Section 2, we introduce the Modified Inverse Exponentiated Exponential Poisson (MIEEP) distribution, elucidating its mathematical and statistical characteristics. Section 3 explores estimation technique namely maximum likelihood (MLE). In Section 4, a dataset on air quality collected in Kathmandu, Nepal in 2021, specifically focusing on PM_{2.5} values, is analyzed. Based on this dataset, we estimate model parameters and demonstrate several criteria for assessing the goodness of fit of the suggested model. In section 5, we conduct time series analysis of air quality data of P_{2.5} in Ratnpark station of Kathmandu, Nepal. Section 6 has closing remarks. The R programming language is used to do numerical computations, contributing to a planned method that guides readers through a clear and well-organized framework. This approach enhances understanding and appreciation of the research methodology and results.

2. Formulation of Modified Inverse Exponentiated Exponential Poisson distribution

In order to create a more versatile probability model, we've incorporated an additional shape parameter, denoted as θ , into (5) of the Cumulative Distribution Function (CDF) of the Inverse Exponentiated Exponential Poisson (IEEP) distribution as described by (Telee & Kumar, 2023)^[34]. This enhanced probability model is referred to as the Modified Inverse Exponentiated Exponential Poisson distribution with three parameters $(\alpha, \beta, \lambda, \theta) > 0$, represented as MIEEP. Consequently, its CDF is expressed by (6).

$$F(x; \beta, \lambda, \alpha, \theta) = 1 - \frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta x^{-1} e^{-\theta x}} \right)^{\alpha} \right\} \right]; x > 0, \lambda > 0, \beta > 0, \alpha > 0, \theta > 0 \quad (6)$$

The associated PDF of equation (6) is included as

$$f(x; \alpha, \beta, \lambda, \theta) = \frac{\alpha \beta \lambda}{(1 - e^{-\lambda})} x^{-2} e^{-\theta x} u [1 + \theta x] (1 - u)^{\alpha-1} \left[\exp \left\{ -\lambda (1 - u)^{\alpha} \right\} \right]; \text{ where, } u = e^{-\beta x^{-1}} e^{-\theta x} \quad (7)$$

Survival function (S(x))

The equation (8) represents the survival function of the proposed model MIEEP.

$$S(x) = \frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta x^{-1}} e^{-\theta x} \right)^{\alpha} \right\} \right]; x > 0, \lambda > 0, \beta > 0, \alpha > 0, \theta > 0 \quad (8)$$

Probability density curves for $\alpha = 0.1657$ and $\beta = 42.9437$ is shown in left side of figure 1. Similarly, pdf curves for $\beta = 42.9437$ and $\lambda = 3$ is shown in right side of the figure 1. The PDF plots show a unimodal distribution that is skewed positively, indicating that values are concentrated around a central point.

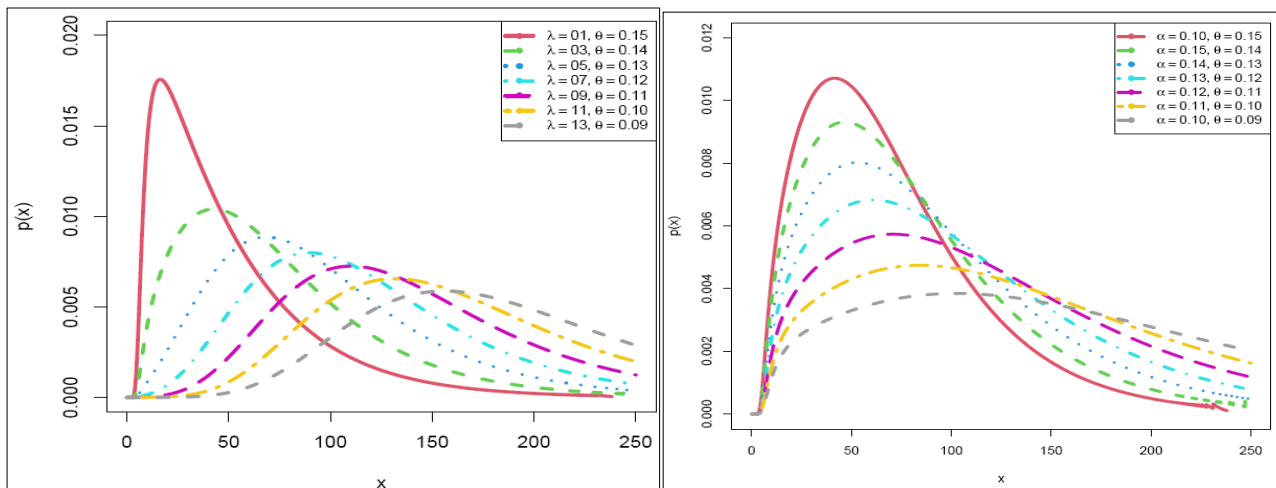


Fig 1: Probability density functions

Hazard rate function (h(x))

The hazard rate represents the rate at which failures occur instantly, as defined by equation (9).

$$h(x) = \alpha \beta \lambda [1 + \theta x] x^{-2} e^{-\theta x} u (1 - u)^{\alpha-1} \left[\exp \left\{ -\lambda (1 - u)^{\alpha} \right\} \right] \left[1 - \exp \left\{ -\lambda (1 - u)^{\alpha} \right\} \right]^{-1} \quad (9)$$

Figure 2 represent the hazard rate curves of the proposed model. Hazard rate curve in left side is for $\beta = 42$, and $\lambda = 0.1$ while curves in right side is for $\beta = 42$, and $\lambda = 3$. The plot of the hazard rate function shows both a rising trend and a distinctive inverted bathtub shape, indicating varying risk levels as time progresses.

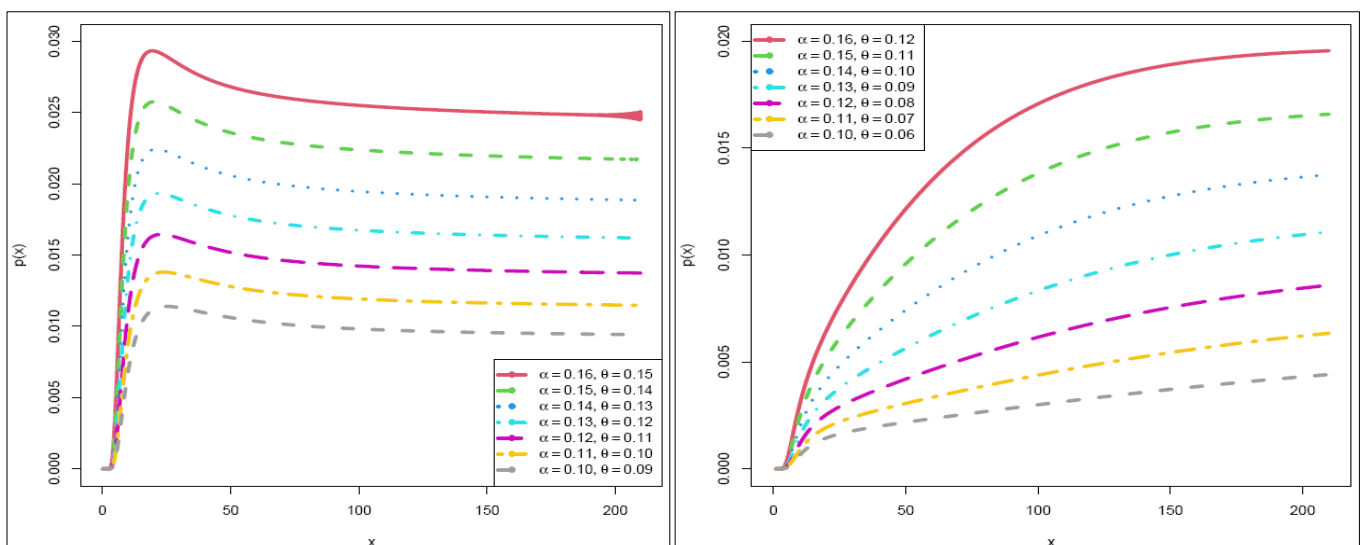


Fig 2: Hazard rate curves

Reversed hazard rate function

The formula (10) defines the model's reversed hazard rate, $h_{rev}(x)$.

$$h_{rev}(x) = \frac{\alpha \beta \lambda [1 + \theta x]}{(1 - e^{-\lambda})} x^{-2} e^{-\theta x} u (1 - u)^{\alpha - 1} \left[\exp \left\{ -\lambda (1 - e^u)^\alpha \right\} \right] \left[1 - \frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda (1 - u)^\alpha \right\} \right] \right]^{-1} \quad (10)$$

Cumulative hazard rate

Expression (11) is the function for cumulative hazard rate

$$H(x) = \ln \left[\frac{1}{(1 - e^{-\lambda})} \left[1 - \exp \left\{ -\lambda \left(1 - e^{-\beta x^{-1} e^{-\theta x}} \right)^\alpha \right\} \right] \right]; x > 0, \lambda > 0, \beta > 0, \alpha > 0, \theta > 0 \quad (11)$$

Quantile function

Quantile function for the MIEEP is defined in (12).

$$e^{-\theta x} + \left(\frac{x}{\beta} \right) \log \left[1 - \left\{ \left(\frac{-1}{\lambda} \right) \log \left\{ 1 - (1 - p) (1 - e^{-\lambda}) \right\} \right\}^{1/\alpha} \right] = 0; \quad 0 \leq p \leq 1 \quad (12)$$

Asymptotic behavior

The asymptotic behavior of the density function may be tested by demonstrating that $\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow \infty} f(x)$.
If the model fits the asymptotic conditions, the modal value will be clearly specified. Considering limits at the endpoints,

$$\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 0} \frac{\alpha \beta \lambda x^{-2} u e^{-\theta x}}{(1 - e^{-\lambda})} [1 + \theta x] (1 - u)^{\alpha - 1} \left[\exp \left\{ -\lambda (1 - u)^\alpha \right\} \right] = 0$$

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

In this case, $\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow \infty} f(x)$, suggesting that the mode of the model will exist.

Skewness

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

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

Kurtosis

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

$$K = \frac{Q(3/8) - Q(0.125) + Q(0.875) - Q(0.625)}{Q(3/4) - Q(1/4)}$$

3. Methods for Estimation of model constants

Method of Maximum likelihood estimation (MLE) approach for estimating the parameters. This estimation method depends on maximizing the log likelihood function of the model. Let's say we have a random sample of 'n' observations from MIEEP, represented by $\underline{x} = (x_1, \dots, x_n)$. The log likelihood function is formulated as follows:

$$\ell(\alpha, \beta, \lambda, \theta | \underline{x}) = n \log(\alpha\beta\lambda) - n \log(1 - e^{-\lambda}) - \lambda \sum_{i=1}^n \left(1 - \exp\left(\frac{-\beta e^{-\theta x_i}}{x_i}\right) \right)^\alpha - \beta \sum_{i=1}^n \left(\frac{e^{-\theta x_i}}{x_i} \right) - \sum_{i=1}^n \left(\theta x_i - \log x_i^{-2} \right) + (\alpha - 1) \sum_{i=1}^n \log \left(1 - \exp\left(\frac{-\beta e^{-\theta x_i}}{x_i}\right) \right) \quad (13)$$

To find the first and second order partial derivatives of the log-likelihood function, differentiate equation (13) with respect to α , β , λ , and θ . By setting the first-order derivatives to zero and solving for the parameters of the proposed model, we can estimate their values. However, in cases where obtaining a solution for the first-order partial derivatives is not feasible, one can employ computer programming to solve the system of nonlinear equations.

4. Real Data Analysis

To assess the model's adaptability, we used the Real-time Air Quality Management System (AQMS) dataset from 2021. Dataset includes Particulate matter (PM) measured at interval of minutes given by (GoN/MoFE, 2023). The following links were used to obtain and acquire real-time air quality datasets for 2021 in Kathmandu, Nepal:

Dataset Link1: http://doenv.gov.np/progressfiles/Final-data-compilation_2021_17-stations-1690193536.xlsx

Dataset Link 2: <http://doenv.gov.np/progress/6/28416357>

In 2016, the Nepalese Department of Environment built 27 Air Quality Monitoring Stations around the country to assess the ground truth of Nepal's air quality conditions using PM_{2.5} measurements. This article examines PM_{2.5} data from Kathmandu, Nepal's Ratnapark Station in 2021. We have also conducted time series analysis of air quality data of P_{2.5} in Ratnapark station of Kathmandu during the years 2017-2021. The optim () function of R software (R Core Team, 2023) was used to evaluate and visualize summary statistics, parameter estimates, model validations, and model comparisons using P_{2.5} data. According to WHO guidelines, PM_{2.5} refers to particulate matter with an aerodynamic diameter of 2.5 μm or less. Air quality data falling outside these ranges can pose health hazards.

Data Analysis for P_{2.5}

Air Quality Standards

National Ambient Air Quality Standards, 2012 (NAAQS) (Lamichhane *et al.*, 2023) ^[20] is legal standards for air quality as Nepalese government. These guidelines provide specific upper limits of some parameters (table1)

Table1: Parameters for quality air and minimum concentration as given by (NAAQS)

SN	Parameters	Unit	Averaging Time	Minimum Concentration
1	TSP	$\mu\text{g}/\text{m}^3$	24-hr	230
2	PM ₁₀	$\mu\text{g}/\text{m}^3$	24-hr	120
3	PM _{2.5}	$\mu\text{g}/\text{m}^3$	24-hr	40
4	Benzene	$\mu\text{g}/\text{m}^3$	Annual	5
5	NO ₂	$\mu\text{g}/\text{m}^3$	Annual	40
			24-hr	80
6	SO ₂	$\mu\text{g}/\text{m}^3$	Annual	50
			24-hr	70
7	Lead	$\mu\text{g}/\text{m}^3$	Annual	0.5
8	Carbon Monoxide	$\mu\text{g}/\text{m}^3$	8-hr	10000
9	Ozon	$\mu\text{g}/\text{m}^3$	8-hr	157

Source: Lamichhane, G.P., Maharjan, N., Pandey, B., K.C., P., Paudel, R., Khanal, S. & Paudel, S.P. (2023) ^[20]. Status of Air Quality in Nepal, Annual Report, 2021. Department of Environment, Kathmandu, Nepal.

Table 2: Summary of the data

Minimum	Q ₁	Q ₂	Mean	Q ₃	Maximum	Sd	Skewness	Kurtosis
7.658	17.423	39.632	47.592	66.261	206.031	35.681	1.391	5.456

Table 2 reveals that the mean P_{2.5} value at Ratnapark Station is higher than the daily median value. Figure 3 depicts the histogram plot (right panel) and the boxplot (left panel).

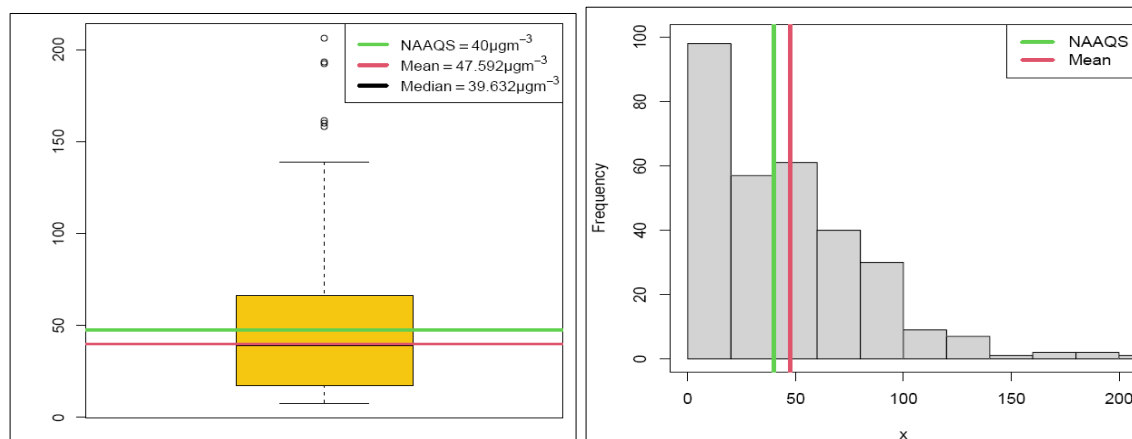
**Fig 3:** Boxplot (left) and histogram (right)

Table 3 displays estimated parameter values using Maximum likelihood estimation method.

Table 3: Estimated parameters values using Maximum likelihood estimation method

Parameters	MLE	SE
α	0.1485662	0.02844442
β	42.243674	3.92273525
λ	0.4320431	0.41941651
θ	0.1522262	0.02908684

Figure 4 presents the histogram alongside the fitted PDF using the MLE estimation approach, in addition to the TTT plot of the suggested model.

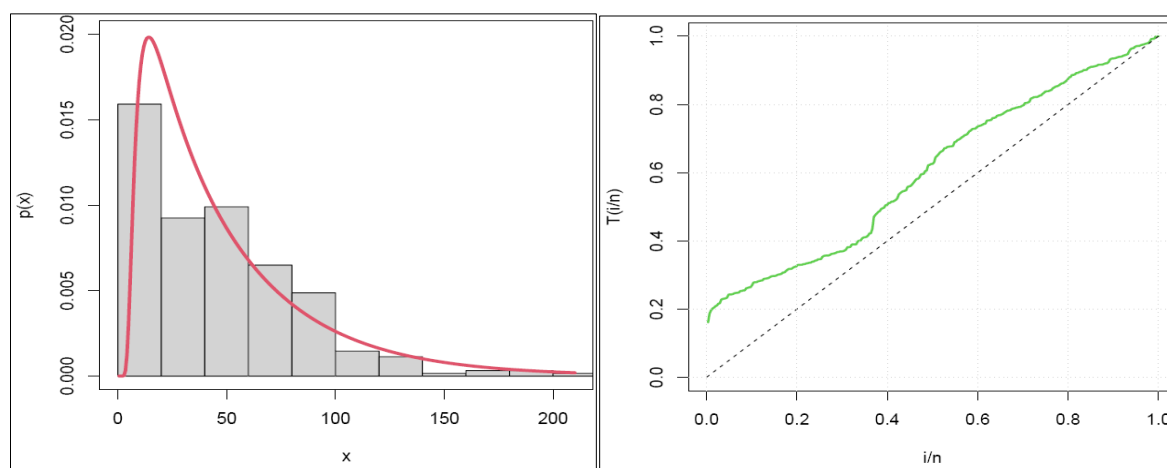
**Fig 4:** TTT plot (Right panel) and Histogram versus pdf (left panel) of MIEEP

Figure 5 shows the empirical CDF curve vs the fitted CDF plot using the suggested model's MLE.

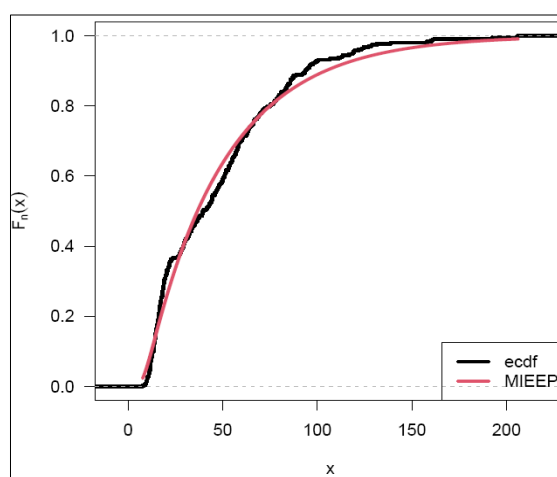
**Fig 5:** Empirical and fitted cdf curve under MLE

Table 4 displays the goodness of fit test statistics values for Anderson-Darling (A^2), Cramer-von Mises (W), and Kolmogorov-Smirnov (KS), as well as the corresponding p-values for the various estimate techniques.

Table4: A^2 , W, and KS statistics with associated p- values

Methods	Critical values	p-value
KS	0.074236	0.067090
W	0.347075	0.100099
A^2	2.475328	0.051073

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

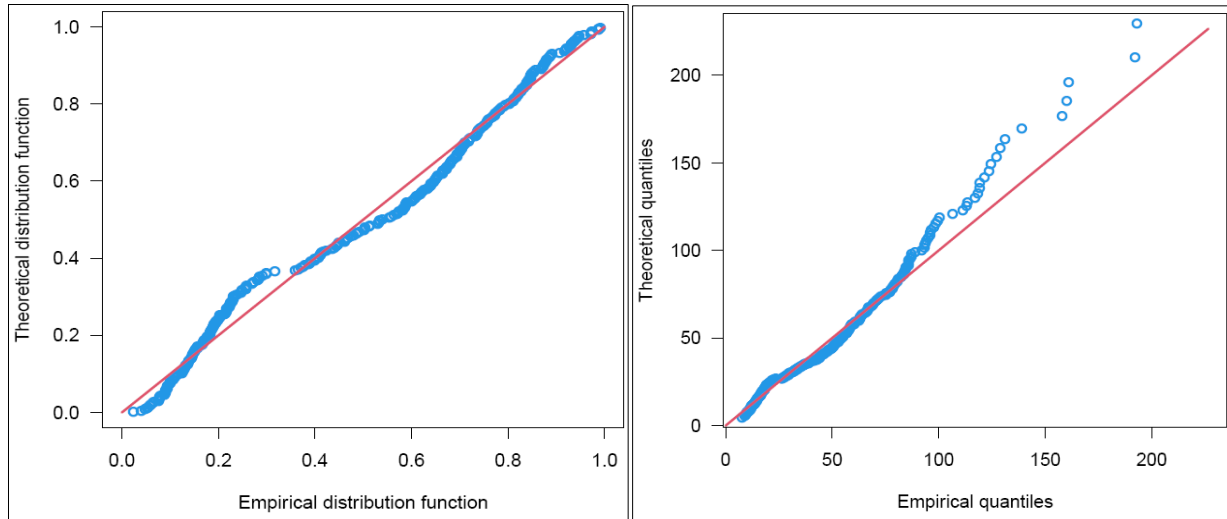


Fig 6: Q & Q plot (right) and P & P plot (left)

We compared five previously published probability models: the Lindley Inverse Weibull (LIW) distributions (Joshi & Kumar, 2020)^[16], the Weibull Extension (WE) distribution (Tang *et al.*, 2003)^[33], the Generalized Weibull Extension (GWE) distribution (Sarhan & Apaloo, 2013)^[30], the Generalized Exponential Extension (GEE) distribution (Lemonte, 2013)^[21] and the Half Logistic Nadarajah Haghighi (HLNHE) distribution (Joshi & Kumar, 2020)^[16].

Table 5 displays the parameters of the models under consideration, which were computed using MLE. Table 5 also shows the SE for each of the distributions analyzed.

Table 5: Estimated parameters and SE for NEKwE, as well as evaluated models.

Model	Alpha	Beta	Theta	Lambda
MIEEP	0.149 (0.028)	42.244 (3.923)	0.152 (0.419)	0.432 (0.029)
GEE	0.467 (0.103)	7.165 (4.376)		0.214(0.221)
LIW	22.991(6.301)	0.493 (0.147)	15.198 (14.470)	-
HLNHE	0.810 (0.207)	1.212 (0.053)	-	0.008 (0.002)
WE	0.003 (0.000)	0.1952 (0.002)	-	0.164 (0.034)
GWE	51.174 (51.174)	51.174 (0.016)	-	51.174 (0.880)

Table 6 shows the information criteria values for MIEEP, as well as the models under consideration. Because the information criterion values for the suggested model are lower than those for the examined model, this suggests that the data fits the suggested model better than the examined models.

Table 6: AIC, BIC, CAIC, Negative Log likelihood (-LL) and HQIC

Models	-LL	AIC	BIC	CAIC	HQIC
MIEEP	1449.189	2906.377	2921.298	2906.509	2912.343
GWE	1608.203	3222.406	3233.596	3222.485	3226.880
GEE	1608.391	3222.783	3233.973	3222.862	3227.257
LIW	1608.719	3223.438	3234.628	3223.517	3227.912
HLNHE	1627.772	3261.545	3272.735	3261.624	3266.019
WE	1636.776	3279.552	3290.742	3279.631	3284.026

We also exhibited the histogram and fitted the density curve for the MIEEP and other models under consideration. Figure 7 shows the model's ecdf as well as the fitted distribution functions for the other models investigated.

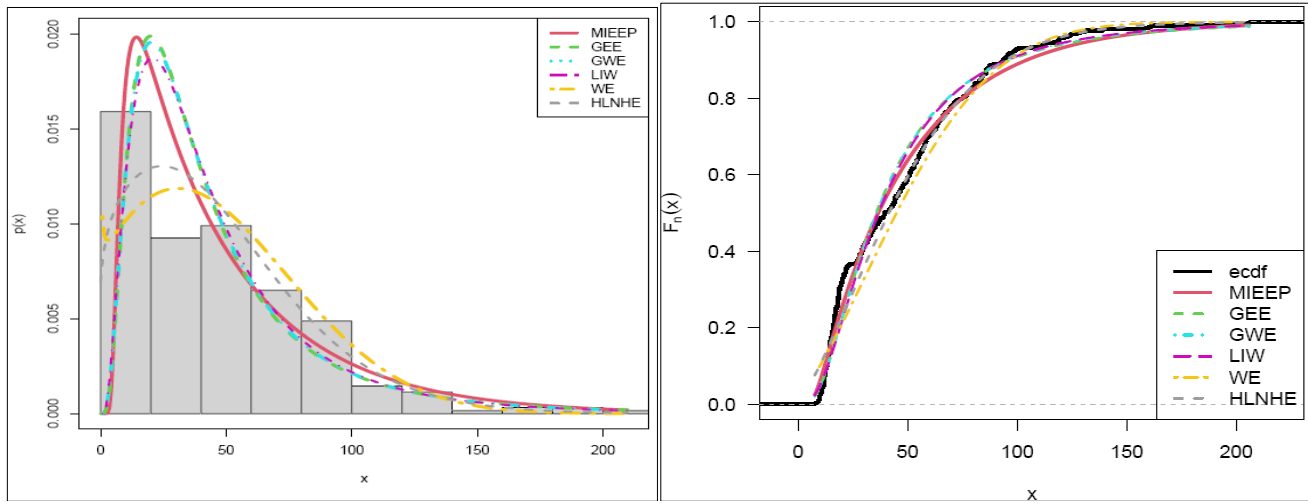


Fig 7: Ecdf with fitted distribution functions of the model(right panel) and Histogram and fitted PDF of models (left panel) .

Table 7 displays the goodness of fit test statistics values for Anderson-Darling (A^2), Kolmogorov-Smirnov (KS) and Cramer-von Mises (W) as well as the related p-values for the proposed and examined models. The recommended model fits data better than the examined models, as indicated by lower test statistics and higher p-values.

Table7: KS, W, and A^2 statistics with corresponding p- values

Models	KS	W	A^2
MEIE	0.0742(0.0671)	0.3471(0.1000)	2.4753(0.0511)
GWE	0.0783(0.0458)	0.5716(0.0263)	3.2072(0.0215)
GEE	0.0779(0.0476)	0.5666(0.0271)	3.1864(0.0221)
LIW	0.0825(0.0303)	0.5605(0.0280)	3.1939 (0.0218)
HLNHE	0.0978(0.0055)	0.5773(0.0254)	4.1993(0.0070)
WE	0.1144(0.0006)	1.0061(0.0024)	6.2948(0.0007)

5. Time series Analysis of air pollution data of PM_{2.5}

Pollution seems to be time series data. It generally depends on the seasons. In this section of the study, we have analyzed the PM_{2.5} data using time series analysis and make some forecasting also. We have divided the whole data into four quarters.

Table 8: Quarterly average PM_{2.5} data

Year	Quarter			
	1	2	3	4
2017	69.873428	46.220952	32.582734	19.560129
2018	72.937583	19.522359	72.110274	20.541654
2019	43.556199	30.879303	59.941249	2.913127
2020	69.319861	21.684052	102.770408	11.120303
2021	75.132993	48.231923	54.721595	16.225978

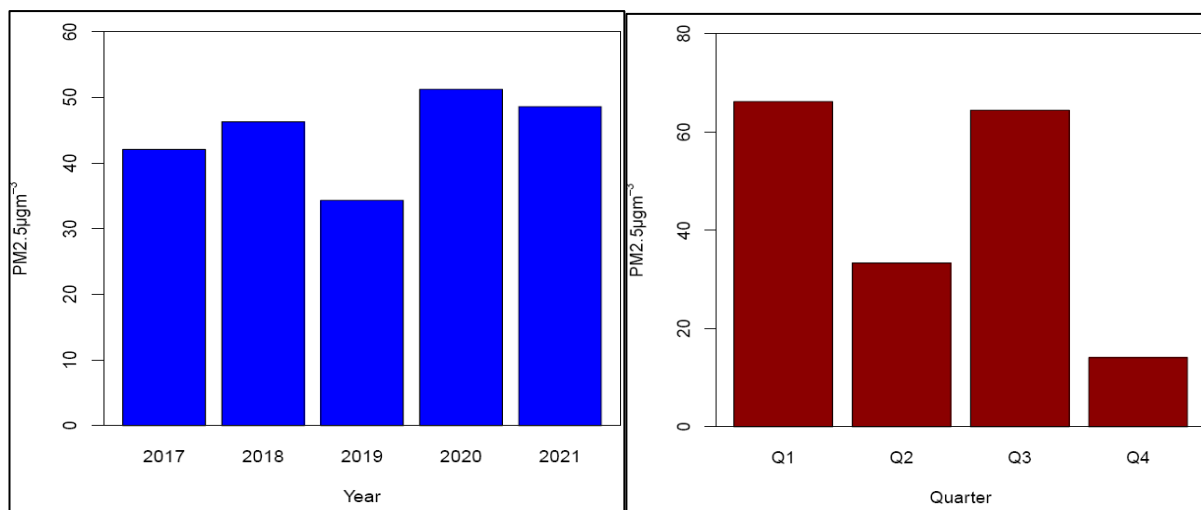


Fig 8: Average PM_{2.5} per year (Left) and average PM_{2.5} per quarter during 2017 to 2021(Right)

Table 8 represents the average $PM_{2.5}$ values for each year from 2017-01-01 to 2021-12-31 for four quarters. The year is split into four quarters: the first (January–March), the second (April–June), the third (July–September), and the fourth (October–December).

Average $PM_{2.5}$ values each year are 42.05931, 46.27797, 34.32247, 51.22366 and 48.57812. Similarly, average values of per quarter during 5 years are 66.16401, 33.30772, 64.4252 and 14.07224 respectively.

Five-year average of $PM_{2.5}$ value is maximum in first quarter equal to 66.164 and minimum equal to 14.072 in fourth quarters as seen in Figure 8. During the five years, 51.224 is maximum value in 2020 whereas 34.322 is minimum value in 2019 (Figure 8).

As data is time series data, so data may have trend, seasonal variation, cyclic variations or, random variations. To study the different variations and trend, we have plotted the graph of observed values in five years shown in left side of figure 9 and observed versus fitted curves in right side.

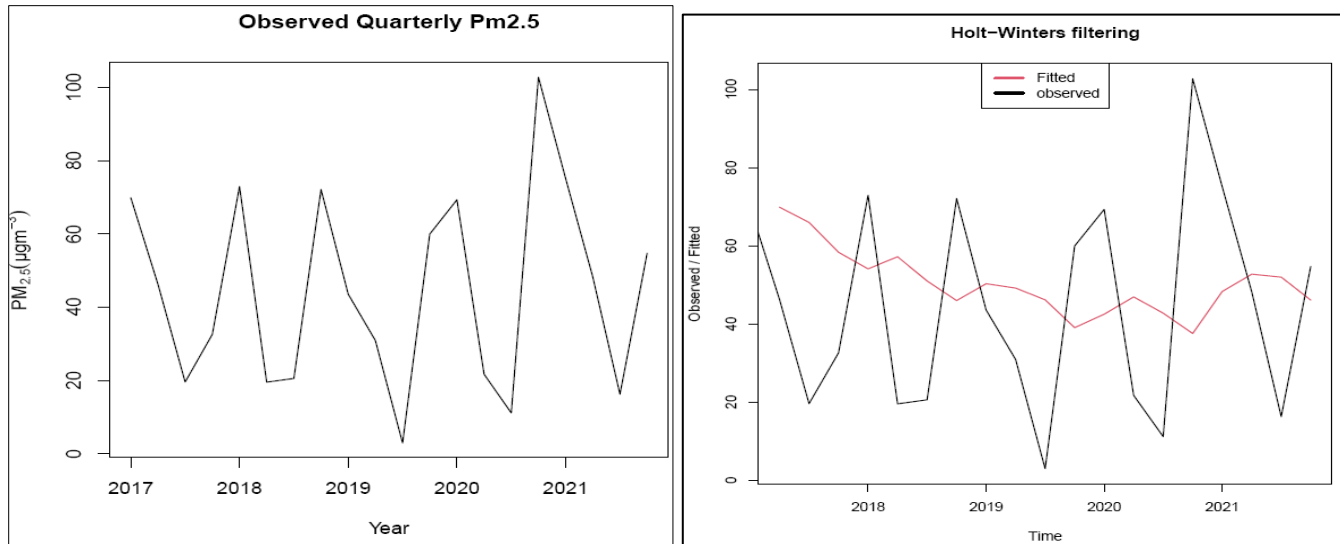


Fig 9: Observed quarterly (left) and observed versus fitted $PM_{2.5}$ (Right)

We can observe from the left time series of $PM_{2.5}$ that there appears to be seasonal fluctuation every year. That is, there is a peak every first quarter, and a trough every fourth quarter. An additive model could be employed to interpret this time series because the seasonal fluctuations seem consistent in size over time, independent of the series' level, and the random fluctuations also appear relatively stable over time. Since data seems to be seasonal, it may contain seasonal, trend and irregular component. We have decomposed these components. The “decompose ()” function in R can be used to estimate the trend and seasonal components. The tables 9,10 and 11 contains estimated values for the seasonal, trend, and irregular components respectively.

The estimated seasonal factors from first to fourth quarter are same across every year. The highest seasonal factor is 22.99534 for quarter 4 and smallest seasonal factor is -30.09482 for third quarter indicating that there seems to be peak in $PM_{2.5}$ in fourth quarter and lowest in third quarter.

Table9: The estimated seasonal factors

	Qtr1	Qtr2	Qtr3	Qtr4
2017	21.42218	-14.32271	-30.09482	22.99534
2018	21.42218	-14.32271	-30.09482	22.99534
2019	21.42218	-14.32271	-30.09482	22.99534
2020	21.42218	-14.32271	-30.09482	22.99534
2021	21.42218	-14.32271	-30.09482	22.99534

Table 10: The estimated trend factors

	Qtr1	Qtr2	Qtr3	Qtr4
2017	NA	NA	42.44233	39.48803
2018	36.27339	41.33703	42.60529	40.35224
2019	39.56829	35.84360	37.54293	39.61398
2020	39.49047	45.87001	51.95030	55.99542
2021	59.95212	54.58422	NA	NA

Table 11: The estimated random factors

Year	Qtr1	Qtr2	Qtr3	Qtr4
2017	NA	NA	7.212615	-29.900634
2018	15.242006	-7.491954	8.031176	8.762692
2019	-17.434278	9.358417	-4.534984	-2.668073
2020	8.407207	-9.863248	-10.735179	23.779643
2021	-6.241308	7.970411	NA	NA

Decomposition and seasonally adjustment:

We have plotted the trend, seasonal, and irregular components of time series given in figure10 (Left).

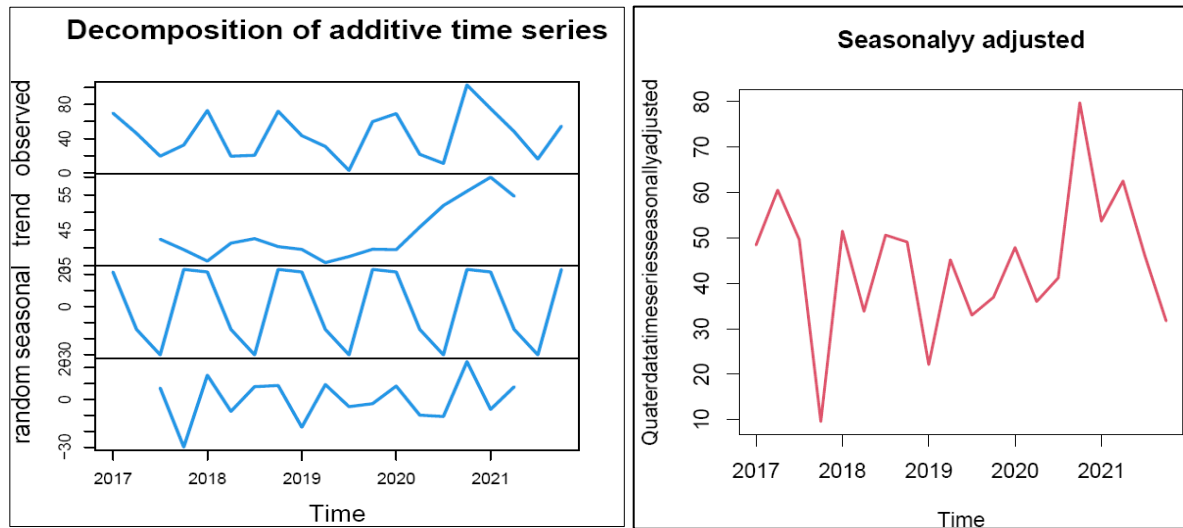


Fig 9: Trend, seasonal, irregular components of time series (Left) and seasonally adjusted time series (Right)

In the plot above, the original time series is displayed at the top, followed by the estimated trend component, the estimated seasonal component, and finally the random component at the bottom. The estimated trend component suggests a little drop from 2018 to 2020 and a rise in 2021.

Here, we have seasonally adjusted the time series by estimating its seasonal components and subtracting these estimates from the original data. This can be done if seasonal time series is described by additive model. This is displayed in figure 9(Right).

Holt-Winters Exponential Smoothing

If your time series shows an additive pattern with a trend (either upward or downward) and seasonal variations, Holt-Winters exponential smoothing is an effective technique for short-term forecasting. This method calculates the present level, trend rate, and seasonal influence of the time series. It involves three key parameters: alpha, beta, and gamma. Alpha determines the smoothing factor for the level, beta controls the smoothing of the trend slope, and gamma adjusts for the seasonal component's variation at the current period. The parameters beta, alpha, and gamma all vary from 0 to 1. When these values are close to 0, they indicate that future value forecasts place relatively little importance on recent observations.

Holt-Winters exponential smoothing with trend and additive seasonal components

Smoothing parameters:

Alpha=0.04881141, beta =0, gamma= 0.4984363

Coefficients: a =35.1244, b = -0.6531, s_1 =32.6662, s_2 = -1.3729, s_3 = -22.9127 and s_4 =31.0102889

The estimated value of alpha is 0.0488, which is quite low. This suggests that the current level estimate heavily relies on both recent observations and some observations from further back in time. Beta's value remains constant at 0.00, indicating that the slope estimate (b) of the trend component remains unchanged throughout the time series, set to its initial value. Gamma is notably high at 0.498, suggesting that the estimate of the seasonal component at the current time point is heavily influenced by the most recent observations. Based on the plot, we observe that the Holt-Winters exponential method effectively predicts the seasonal peaks, typically occurring in the fourth quarter each year. The blue line in Figure 10 reflects the original time series, while the red line on the left side indicates the predicted time series.

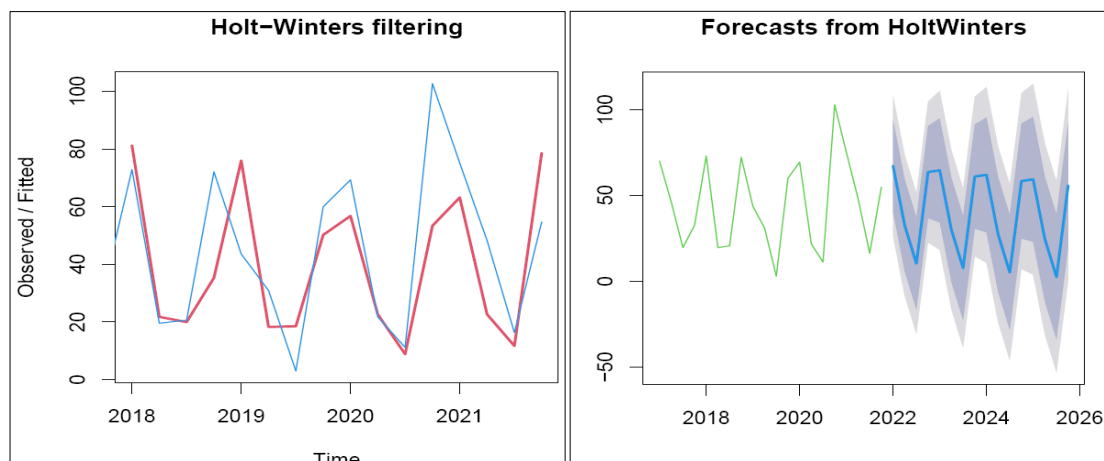


Fig 10: Exponential smoothing (Left) forecasted graph (Right)

We have also forecasted for future times not in our time series data. Original data is from first quarter 2017 to fourth quarter 2021, all together 20 quarters. We have forecasted for first quarter 2022 to fourth quarter 2030 (36 more quarters). Forecasted values are tabulated below in table12.

Table 12: Forecasted values of PM_{2.5} data.

Year/Quatr	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2022 Q1	67.137714	40.3558577	93.91957	26.178399138	108.09703
2022 Q2	32.445379	5.6316370	59.25912	-8.562700856	73.45346
2022 Q3	10.252596	-16.5929942	37.09819	-30.804191170	51.30938
2022 Q4	63.522501	36.6451004	90.39990	22.417064194	104.62794
2023 Q1	64.525495	34.2182855	94.83270	18.174618060	110.87637
2023 Q2	29.833160	-0.5022299	60.16855	-16.560815193	76.2271
2023 Q3	7.640376	-22.7231675	38.00392	-38.796656688	54.07741
2023 Q4	60.910281	30.5186090	91.30195	14.430229670	107.39033
2024 Q1	61.913275	28.4500713	95.37648	10.735721105	113.09083
2024 Q2	27.220940	-6.2677886	60.70967	-23.995650721	78.43753
2024 Q3	5.028157	-28.4860773	38.54239	-46.227441058	56.28375
2024 Q4	58.298061	24.7583414	91.83778	7.003486232	109.59264
2025 Q1	59.301056	22.9548714	95.64724	3.714363738	114.88775
2025 Q2	24.608720	-11.7609652	60.97841	-31.013913785	80.23135
2025 Q3	2.415937	-33.9772349	38.80911	-53.242616338	58.07449
2025 Q4	55.685842	19.2691986	92.10248	-0.008607762	111.38029

The forecasts are depicted by a blue line, with shaded areas in violet and black indicating 80% and 95% prediction intervals, respectively. Additionally, we assessed the potential improvement of the predictive model by investigating non-zero autocorrelations in the forecast errors across lags 1 to 12. This evaluation involved generating a correlogram and executing the Ljung-Box test.

The correlogram shows that the autocorrelations of the forecast errors within the sample do not exceed the significance limits for lags 1-12. Moreover, the Ljung-Box test yields a p-value of 0.6069, indicating minimal evidence of nonzero autocorrelations at lags 1-12.

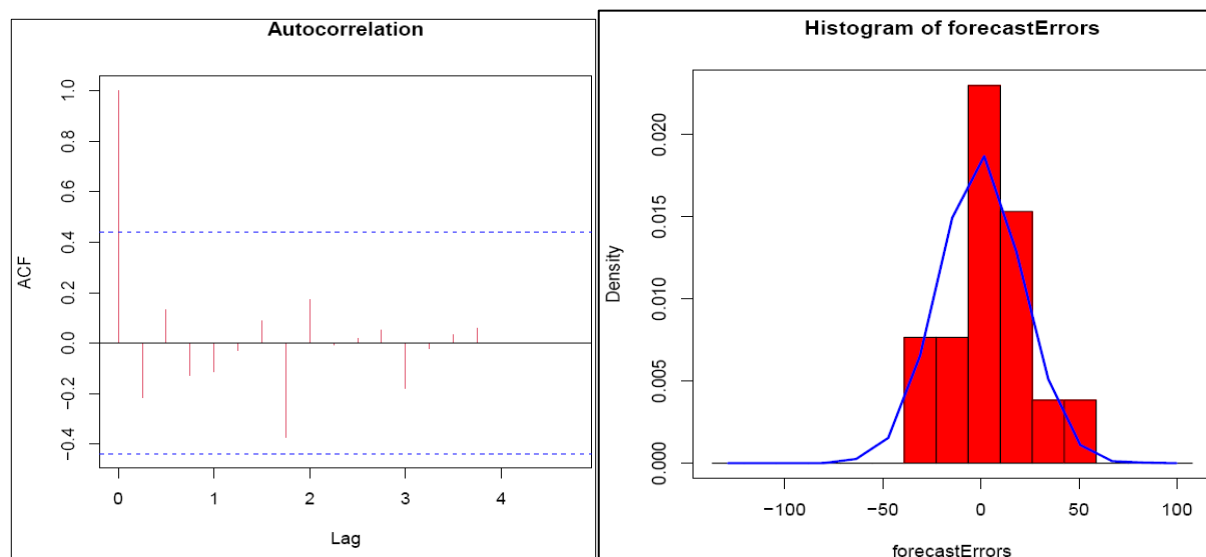


Fig 11: Autocorrelation function and Histogram of forecast errors

Box-Ljung test

Data: Quaterdatatimeseriesforecasts2\$residuals

X-squared = 10.104, df = 12, p-value = 0.6069

We also examined whether the forecast errors exhibit consistent variance over time and follow a normal distribution with a non-zero mean. To assess this, we plotted the forecast errors and a histogram on the right side of Figure 11. The analysis indicates that the predicted errors are most likely normally distributed with a mean of zero. Furthermore, the Ljung-Box test suggests minimal evidence of non-zero autocorrelation in the forecast errors over the sample period. This reinforces the observation that the forecast error distribution approximates normality with a mean of zero. Therefore, these findings imply that the simple exponential smoothing method offers a suitable predictive model for the PM_{2.5} data spanning from 2017 to 2021.

6. Conclusion

In this study, a novel and versatile probability distribution known as the Modified Inverse Exponentiated Exponential Poisson distribution has been introduced. We thoroughly examined various statistical properties of this innovative model. To evaluate the appropriateness of our model, we analyze a real dataset that includes P_{2.5} air quality measurements obtained from the Ratnapark

station in Kathmandu, Nepal, over the course of 2021. We performed data analysis to estimate parameters, validate models, and compare models for $P_{2.5}$ data. However, concentrations of $P_{2.5}$ were at their lowest during the monsoon season and peaked during the winter season, indicating seasonal fluctuations.

We used a variety of diagnostic methods to validate our model, including P-P and Q-Q graphs. In addition, we evaluated metrics such as log-likelihood values, the Corrected Akaike information criterion, Hannan-Quinn information criteria, Akaike information criterion and Bayesian information criterion. To assess the model's suitability and accuracy, we performed Anderson-Darling, Cramer-Von Mises, and Kolmogorov-Smirnov tests. We have also conducted the time series analysis of air pollution data of $P_{2.5}$ during years 2017-2021 in Ratnapark of Kathmandu, Nepal. Based on our findings, the air quality in Kathmandu, Nepal was consistently poor, except during the COVID-19 pandemic. Additionally, we determined that the simple exponential smoothing method effectively predicts $PM_{2.5}$ levels from 2017 to 2021.

7. References

1. Adamidis A, Loukas S. A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters*. 1998;39:35-42. [https://doi.org/10.1016/S0167-7152\(98\)00012-1](https://doi.org/10.1016/S0167-7152(98)00012-1)
2. Al-saiary ZA, Bakoban RA, Al-zahrani AA. Characterizations of the Beta Kumaraswamy Exponential Distribution. *Mathematics*. 2019;8(1):23. <https://doi.org/10.3390/math8010023>
3. Asgharzadeh A, Bakouch HS, Esmaeili L. Pareto Poisson-Lindley distribution with applications. *Journal of Applied Statistics*. 2013;40(8):1717-1734.
4. Bakouch HS, Ristić MM, Asgharzadeh A, Esmaily L, Al-Zahrani BM. An exponentiated exponential binomial distribution with application. *Statistics and Probability Letters*. 2012;82(6):1067-1081.
5. 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>
6. Barreto-Souza W, Cribari-Neto F. A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*. 2009;79(24):2493-2500. <https://doi.org/10.1016/j.spl.2009.09.003>
7. 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>
8. 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>
9. 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
10. Chaudhary AK, Teele 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>
11. 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>
12. 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
13. Gupta RD, Kundu D. Theory and methods: Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*. 1999;41(2):173-188. <https://doi.org/10.1111/1467-842X.00072>
14. Gupta RD, Kundu D. Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*. 2001;43(1):117-130.
15. Jayakumar K, Sankaran KK. Exponential intervened Poisson distribution. *Communications in Statistics-Theory and Methods*. 2021;50(13):3063-3093. <https://doi.org/10.1080/03610926.2019.1682161>
16. Joshi RK, Kumar V. Half Logistic NHE Distribution: Properties and Application. *International Journal for Research in Applied Science and Engineering Technology*. 2020;8(IX):742-753. <https://doi.org/10.22214/ijraset.2020.31557>
17. Joshi RK, Kumar V. Lindley inverse Weibull distribution: Theory and Applications. *Bulletin of Mathematics and Statistics Research*. 2020;8(3):32-46. <http://www.bomsr.com/>
18. Joshi RK, Kumar V. New Exponential Extension Poisson Distribution: Properties and Applications. *International Journal of Mathematics and Statistics Invention*. 2021;8(9):35-45.
19. Kus C. A new lifetime distribution. *Computational Statistics and Data Analysis*. 2007;51(9):4497-4509. <https://doi.org/10.1016/j.csda.2006.07.017>
20. 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.
21. 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>
22. 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>
23. Mahmoudi E, Sepahdar A. Exponentiated Weibull-Poisson distribution: Model, properties and applications. *Mathematics and Computers in Simulation*. 2013;92:76-97. <https://doi.org/10.1016/j.matcom.2013.05.005>
24. 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>

25. Maya R, Chesneau C, Krishna A, Irshad MR. Poisson Extended Exponential Distribution with Associated INAR (1) Process and Applications. *Stats*. 2022;5(3):755-772. <https://doi.org/10.3390/stats5030044>
26. Moors JJA. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1988;37(1):25-32. <https://doi.org/10.2307/2348376>
27. Muhammad M. Poisson-odd generalized exponential family of distributions: Theory and applications. *Haceteppe Journal of Mathematics and Statistics*. 2018;47(6):1652-1670.
28. R Core Team. R: A Language and environment for statistical computing. (Version 4.1) [Computer software]; c2023. Retrieved from <https://cran.r-project.org> (R packages retrieved from CRAN snapshot 2023-04-07).
29. Ristic MM, Nadarajah S. A new lifetime distribution. *Journal of Statistical Computation and Simulation*. 2014;84(1):135-150. <https://doi.org/10.1080/00949655.2012.697163>
30. Sarhan AM, Apaloo J. Exponentiated modified Weibull extension distribution. *Reliability Engineering and System Safety*. 2013;112:137-144. <https://doi.org/10.1016/j.res.2012.10.013>
31. Shrestha SL. Analysis of ambient particulate air pollution and health in NEPAL. *Journal of Global Ecology and Environment*. 2021;12(4):14-29.
32. Sousa-Ferreira I, Abreu AM, Rocha C. The Extended Chen-Poisson Lifetime Distribution. *REVSTAT-Statistical Journal*. 2023;21(2):173-196. <https://orcid.org/0000-0001-5526-3594>
33. Tang Y, Xie M, Goh TN. Statistical analysis of a Weibull extension model. *Communications in Statistics-Theory and Methods*. 2003;32(5):913-928. <https://doi.org/10.1081/STA-120019952>
34. Telee LBS, Kumar V. Inverse Exponentiated Exponential Poisson Distributions with theory and applications. *International Journal of Engineering Science Technologies*. 2023;7(5):17-36. <https://doi.org/10.29121/ijoest.v7.i5.2023.484>
35. Wahid A, Al-Zahrani BM, Mansoor M, Merovci F. The Gamma-Weibull Poisson distribution: Properties, regression model and application. *Journal of Applied Statistics*. 2018;45(1):47-66. <https://doi.org/10.1080/02664763.2017.1298131>
36. 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. <https://doi.org/10.3390/math10122055>
37. Xie M, Tang Y, Goh TN. A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*. 2002;76(3):279-285. [https://doi.org/10.1016/S0951-8320\(02\)00036-4](https://doi.org/10.1016/S0951-8320(02)00036-4)
38. Zarei H, Madadi MA, Khorrami M, Zeal G. Improved Inference on the Parameters of Pareto-Poisson Exponential Distribution Based on Lower Record Data. *Scientia Iranica*. 2020;27(6):3456-3469. <https://doi.org/10.24200/sci.2020.54394.3748>