Application of the Parametric Bootstrap Method for Confidence Interval Estimation and Statistical Analysis of PM2.5 in Bangkok

BOONYARIT CHOOPRADIT*, RUJAPA PAITOON, NATTAWADEE SRINUAN, SATITA KWANKAEW
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathumthani,
THAILAND

**Corresponding Author*

Abstract: - Research in epidemiology and health science indicates that exposure to particles with an aerodynamic diameter of less than 2.5 µm (PM2.5) causes harmful health consequences. Probability density functions (pdf) are utilized to analyze the distribution of pollutant data and study the occurrence of high-concentration occurrences. In this study, PM2.5 concentrations (in $\mu g/m^3$) were recorded daily from January 2011 to December 2022 at 12 air quality monitoring locations in Bangkok. The study utilized two-parameter distributions such as gamma, inverse Gaussian, lognormal, log-logistic, Weibull, and Pearson type V to identify the most suitable statistical distribution model for PM2.5 in Bangkok. The Anderson-Darling test result indicates that the inverse Gaussian and Pearson type V distributions are the most appropriate probability density functions for the daily average PM2.5 concentration at stations in Bangkok. The projected 98th percentile of daily PM2.5 levels at two locations is higher than the 24-hour threshold for daily PM2.5 concentrations in Thailand, posing significant health risks. Additionally, the two parametric bootstrap methods used to estimate confidence intervals for the median, namely percentile bootstrap and simple bootstrap, indicate that two stations have poor air quality for those with sensitive health conditions.

Key-Words: - PM2.5, Air pollution, Statistical analysis, Distribution, Parametric bootstrap, Bangkok

Received: June 25, 2023. Revised: March 19, 2024. Accepted: April 21, 2024. Published: May 22, 2024.

1 Introduction

Numerous epidemiological and toxicological studies have consistently identified ambient fine particulate matter (PM2.5) as a significant and detrimental factor for human health. PM2.5 refers to particles with a diameter of less than 2.5 µm, [1]. In 2019, air pollution was classified as the fourth-most significant risk factor in terms of its contribution to premature mortality worldwide. Furthermore, in terms of worldwide relevance, only high blood pressure, tobacco consumption, and insufficient eating habits exceeded it as contributing factors. Countries situated in Asia, Africa, and the Middle East persistently grapple with the most substantial levels of atmospheric pollutant particle matter. Based on the Health Effects Institute's 2020 study, this harmful pattern underscores the ongoing and substantial problems that these specific regions have about air quality and the associated public health risks, [2]. An estimation study by [3] found a considerable number of deaths, more than 5500, linked to PM2.5 pollution in the environment.

Thailand is rated 57th in the world for air quality according to the 2022 World Air Quality Report by [4], implying its significance as a nation with environmental pollution issues. In 2022, Thailand had an average concentration of PM2.5 particles lowering to $18.1\,\mu g/m^3$, meaning its air quality increased. This indicates a sharp drop of 10.4% from the amounts mentioned in 2021. The provinces of Khon Kaen, Mae Hong Son, Chiang Mai, Bangkok, Nonthaburi, and Nakhon Ratchasima faced the highest density of PM2.5 concentrations according to the geographical breakdown of air pollution in Thailand. The information shows that there are often geographical differences in the quality of the air, with the provinces that are emphasized having higher pollution readings. In addition, Bangkok, the capital of Thailand, ranks 52nd in the world among capital cities in a global study of capital cities. Indicating the trials faced by a city when trying to resolve issues concerning air quality at the urban level, Bangkok conveyed an average PM2.5 concentration of 18 $\mu g/m^3$ in 2022. According to [5], we have determined that the average concentration of PM2.5 during a 24-hour period >75 $\mu g/m^3$ indicates that the air quality level is harmful to people's health.

The study of statistical distributions is important in the field of statistics as it reveals patterns of data, [6], [7]. The statistical distribution model is utilized to analyze the distribution of air pollutant data and assess the occurrence of high-concentration occurrences. By choosing a suitable statistical distribution for air pollutants, the mean or median concentration may be reliably predicted, [8], [9], [10]. For this reason, a variety of statistical distributions have been used in literature to describe concentrations of PM2.5 distributions including lognormal (LN), gamma, inverse Gaussian (IG), log-logistic (LL), Weibull, and Pearson type V or Inverse gamma (Pearson V) distributions, [9], [10], [11].

A confidence interval (CI) is a statistical range of values that is highly probable to encompass the actual value of the population parameter of interest. The CI for a parameter of interest is calculated by removing or adding the product of the standard error and a critical value. The computation assumes that the estimate of the parameter conforms to an approximately normal distribution. Regrettably, when the assumption of normality is broken or the maximum likelihood estimator cannot be easily determined, it may be challenging to calculate these CIs. In such cases, an additional method like the bootstrap method can be employed, [12], [13], [14], [15], [16].

Familiarity with the statistical distribution is a crucial element in comprehending the statistical characteristics of PM2.5 levels in Bangkok. Prior studies have mostly examined the statistical characteristics of PM2.5 levels and determined the suitable distribution for PM2.5 values, [9], [10], [11]. Therefore, this study analyzed the daily PM2.5 values recorded at twelve air quality monitoring sites in Bangkok. These data were analyzed using six two-parameter statistical distribution models to determine the most suitable distribution for capturing the daily PM2.5 values. Furthermore, utilize a suitable distribution for estimating median confidence intervals. The median values across all stations were calculated with parametric bootstrap confidence intervals. This is an advantage of this study if there is no closed-form equation for the median. The subsequent sections of this work are structured in the following manner: Section 2 contains a detailed explanation of the materials and techniques used in this study. Section 3 shows the results. Finally, Section 4 includes the conclusion and provides recommendations for future research.

2 Methods and Materials

2.1 Data

The current investigation acquired historical daily concentrations of PM2.5 data (in $\mu g/m^3$) from January 2011 to December 2022 for the 12 air quality monitoring sites in Bangkok. This data was collected from the Pollution Control Department under the Ministry of Natural Resources and Environment, [17]. The station codes are displayed below: X02T - Bansomdejchaopraya Rajabhat University, X03T - Along Kanchanaphisek Road, X05T - Meteorological Department (Bang Na), X10T - Klong Chan Housing Community, X11T - Huai Khwang Housing Community Stadium, X12T - Nonsi Witthaya School, X50T - Chulalongkorn Hospital, X52T - Metropolitan Electricity Authority Thonburi Substation, X53T - Chokchai Metropolitan Police Station, X54T - Din Daeng Housing Community, X59T - The Government Public Relations Department, X61T - Bodindecha (Sing Singhaseni) School.

2.2 Distributions

This study primarily examines six two-parameter distributions often employed to model daily concentrations of PM2.5 data: gamma, inverse Gaussian (IG), lognormal (LN), log-logistic (LL), Weibull, and Pearson type V or Inverse gamma (Pearson V) distributions. Table 1 summarizes the probability density function (PDF), cumulative distribution function (CDF), and distribution parameters of these six distributions, [18], [19], [20], [21], [22].

2.3 Maximum Likelihood Estimation

The specific characteristics of a theoretical distribution are contingent upon the precise values of its parameters. The most suitable parameter values for the distributions were determined through the utilization of the maximum likelihood estimation (MLE) method. Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a PDF of each x_i is $f(x_i;\theta_1,\theta_2,\ldots,\theta_k)$ where θ_k is an unknown k-parameters. The likelihood function $\mathrm{L}(\theta)$ associated with this random sample is defined as a function of the unknown parameters in the following manner:

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\ldots,\theta_k)$$

where $x_1, x_2, \ldots, x_n$ are the independent observations from a random sample. Given the differentiability of the likelihood function at $\theta_1, \theta_2, \ldots, \theta_k$, the estimation of the parameters through the maximum likelihood method involves computing partial derivatives of concerning each parameter and subsequently solving the resulting set of *k* equations to find their zeros. Typically, computations involve utilizing the logarithm of the likelihood function, given that it is a strictly increasing function. Consequently, the same parameter values will optimize both the likelihood function and the log-likelihood function. Thus, the MLE of $\theta_1, \theta_2, \ldots, \theta_k$ is a solution of

$$\frac{\partial \ln \mathrm{L}(\theta)}{\partial \theta_k} = 0 .$$

The MLE for $\theta_1, \theta_2, \ldots, \theta_k$ can be derived by solving the resulting equations simultaneously through a numerical procedure, such as the Newton-Raphson method. In this study, the MLE estimates of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are obtained using the "fitdist" function from the "fitdistrplus" package in the R software suite, [23].

2.4 Assessment of Goodness-of-fit Test

The aforementioned six functions were compared using various statistical approaches, such as the Kolmogorov-Smirnov, Anderson-Darling (AD), and Chi-Square goodness-of-fit tests. Nevertheless, the empirical data provided by [24], indicates that the AD goodness-of-fit test statistic, [25], is superior to other tests when it comes to evaluating the suitability of a positively skewed distribution. Hence, it is advisable to utilize the AD test to assess the adequacy of fit in this particular study, given this discovery. The test statistic for the AD test may be denoted by the following mathematical expression:

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln\left(F\left(x_i\right)\right) + \ln\left(1 - F\left(x_{n+1-i}\right)\right) \right],$$

where $F(.)$ is the expected CDF, x_i are the data must be put in order $\left(x_1 < \ldots < x_n\right)$, and n is the sample size. A lower AD statistic result indicates the appropriateness of the statistical distribution being examined, [7].

Table 1. Statistical properties of daily concentrations of PM2.5 distribution functions

Distribution	PDF	CDF	Parameter
gamma	$f(x;\alpha,\beta) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp(-x/\beta)$, $x \geq 0$	$\mathrm{F}(x;\alpha,\beta) = \frac{\Gamma_{x/\beta}(\alpha)}{\Gamma(\alpha)}$	shape parameter $\alpha > 0$ scale parameter $\beta > 0$
IG	$f(x;\mu,\theta) = \left(\frac{\theta}{2\pi x^3}\right)^{1/2}\exp\left(-\frac{\theta(x-\mu)^2}{2\mu^2 x}\right)$, $x > 0$	$\mathrm{F}(x;\mu,\theta) = \Phi\left[\left(\frac{\theta}{x}\right)^{1/2}\left(\frac{x}{\mu}-1\right)\right] + \exp\left(\frac{2\theta}{\mu}\right)\Phi\left[-\left(\frac{\theta}{x}\right)^{1/2}\left(\frac{x}{\mu}+1\right)\right]$	shape parameter $\theta > 0$ mean $\mu > 0$
LN	$f(x;\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$, $x > 0$	$\mathrm{F}(x;\mu,\sigma^2) = \Phi\left(\frac{\ln x-\mu}{\sigma}\right)$	location parameter μ, $-\infty < \mu < \infty$ scale parameter $\sigma > 0$
LL	$f(x;\gamma,\theta) = \frac{\gamma(x/\theta)^{\gamma}}{x\left[1+(x/\theta)^{\gamma}\right]^2}$, $x > 0$	$\mathrm{F}(x;\gamma,\theta) = \frac{(x/\theta)^{\gamma}}{1+(x/\theta)^{\gamma}}$	scale parameter $\theta > 0$ shape parameter $\gamma > 0$
Weibull	$f(x;\beta,\alpha) = \left(\frac{\beta x^{\beta-1}}{\alpha^{\beta}}\right)\exp\left[-(x/\alpha)^{\beta}\right]$, $x \geq 0$	$\mathrm{F}(x;\beta,\alpha) = 1-\exp\left[-(x/\alpha)^{\beta}\right]$	shape parameter $\beta > 0$ scale parameter $\alpha > 0$
Pearson V	$f(x;\alpha,\beta) = \frac{\exp(-\beta/x)}{\beta\Gamma(\alpha)(x/\beta)^{\alpha+1}}$, $x > 0$	$\mathrm{F}(x;\alpha,\beta) = 1-\Gamma(\alpha;\beta/x)$	shape parameter $\alpha > 0$ scale parameter $\beta > 0$

2.5 Parametric Bootstrap Method for Estimating Confidence Interval for the Median

When dealing with skewed distributions, it is important to note that the arithmetic mean, sometimes known as the "mean," tends to be located more toward the tail of the distribution compared to the median. Therefore, the median is often preferred as a measure of central tendency because the mean does not always correspond to the center location in the distribution, [26]. The determination of the median (M) of a distribution depends on the specific characteristics that define that distribution. To calculate the median, one must use the cumulative distribution function (CDF) of the particular distribution being considered, denoted as CDF(x). To get the median precisely, we need to solve the equation CDF(M) = 0.5 using numerical techniques. Regrettably, the Pearson type V distribution lacks a straightforward mathematical equation to calculate its median, unlike several other distributions. Nevertheless, it is possible to approximate the median by utilizing the quantile function of the Pearson V distribution. Constructing a confidence interval for the median of the Pearson V distribution requires the use of statistical estimating techniques, such as the bootstrap method, because there is no closed-form equation for the median. The bootstrap methodology was initially described by [27], who provided a comprehensive explanation of the fundamental principles underlying the basic bootstrap approach. In [16] the authors presented a comprehensive summary of the various bootstrap methods. The authors classified and contrasted several bootstrap approaches, demonstrating their respective benefits. A detailed analysis, especially comparing bootstrap confidence interval approaches, can be found in the study conducted by [15]. Their study examined the characteristics of several bootstrap confidence interval methods. Further analysis of the options for applying bootstrap confidence intervals is provided by [14]. This resource offers practical advice on choosing and implementing suitable methods for bootstrap confidence intervals.

2.5.1 Parametric Percentile Bootstrap (PB) Method

The procedure for constructing a parametric PB confidence interval for the median of any distribution may be summarized as follows:

1) Fit a parametric distribution to the original sample data. Estimate the distribution parameters.

2) Generate B bootstrap samples by sampling randomly with replacement from the fitted parametric distribution. The bootstrap sample size equals the original sample size. This study uses B = 100,000.

3) Compute the median of each bootstrap sample. This gives B estimates of the median.

4) Sort the B bootstrap median estimates in ascending order.

5) The $(\alpha/2)B^{\text{th}}$ and $(1-\alpha/2)B^{\text{th}}$ values in the sorted bootstrap medians provide the lower and upper confidence limits of a $100(1-\alpha)\%$ CI for the true median.

6) Typically α = 0.05 is used for a 95% CI. So the 2.5^{th} and 97.5^{th} percentile bootstrap medians give the 95% CI endpoints.

2.5.2 Parametric Simple Bootstrap (SB) Method

The procedure for constructing a parametric SB confidence interval for the median of any distribution may be summarized as follows:

1) Fit a parametric distribution to the original sample data. Estimate the distribution parameters. Find the median (m) from the original sample data.

2) Generate B bootstrap samples by sampling randomly with replacement from the fitted parametric distribution. The bootstrap sample size equals the original sample size. This study uses B = 100,000.

3) Compute the median of each bootstrap sample. This gives B estimates of the median.

4) Sort the B bootstrap median estimates in ascending order.

5) The $(\alpha/2)B^{\text{th}}$ and $(1-\alpha/2)B^{\text{th}}$ values in the sorted bootstrap medians represent the lower (L) and upper (U) limits.

6) The $2m$ - U and $2m$ - L values provide the lower and upper confidence limits of a $100(1-\alpha)\%$ CI for the true median.

3 Results and Discussion

Section 3.1 presents a summary of the statistical properties of the daily PM2.5 concentrations measured at the 12 chosen stations in Bangkok. Subsequently, in Section 3.2, the suitable distributions and estimated parameters for daily PM2.5 levels at each station are determined. Calculations are conducted to determine the percentiles and exceedance probability for each station indicated in Section 3.3. Section 3.4 calculates confidence intervals for the median at

DOI: 10.37394/232015.2024.20.22

each station using the parametric bootstrap technique.

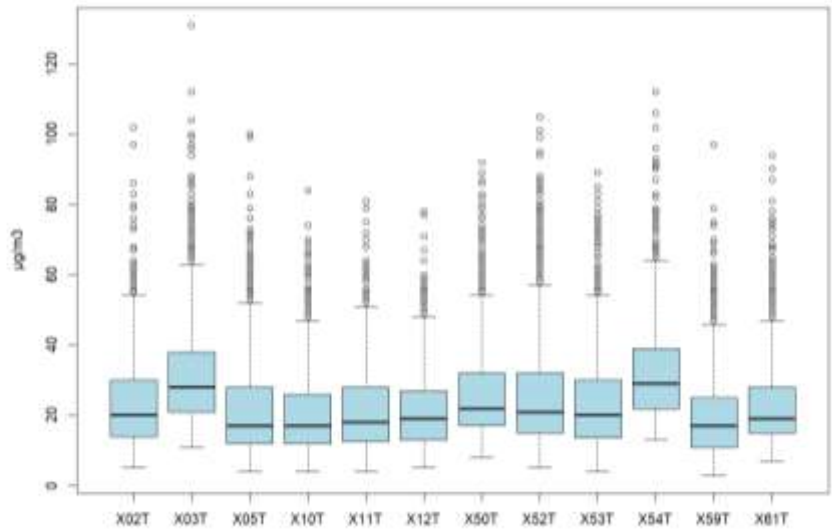

Fig. 1: Boxplots displaying the daily measurements of PM2.5 levels at the specified station in Bangkok

Table 2. Descriptive summary of daily PM2.5 levels (in $\mu g/m^3$) at the specified station in Bangkok

Station	n^a	Mean	Median	Mode	SD[b]	Min.	Max.	Q_1	Q_2	Q_3	Skewness	Kurtosis
X02T	1221	23.91	20.00	14.00	13.45	5	102	14.00	20.00	30.00	1.53	3.10
X03T	1524	31.63	28.00	21.00	14.71	11	131	21.00	28.00	38.00	1.77	4.56
X05T	2117	21.65	17.00	13.00	13.83	4	100	12.00	17.00	28.00	1.47	2.42
X10T	1477	20.55	17.00	11.00	11.86	4	84	12.00	17.00	26.00	1.43	2.29
X11T	1528	21.84	18.00	10.00	12.29	4	81	12.25	18.00	28.00	1.34	1.88
X12T	1217	21.41	19.00	12.00	11.12	5	78	13.00	19.00	27.00	1.36	2.16
X50T	2007	26.04	22.00	16.00	12.63	8	92	17.00	22.00	32.00	1.55	2.84
X52T	2160	25.60	21.00	15.00	15.03	5	105	15.00	21.00	32.00	1.63	3.12
X53T	1972	23.58	20.00	14.00	13.79	4	89	13.25	20.00	30.00	1.43	2.31
X54T	1607	32.56	29.00	22.00	13.78	13	112	22.00	29.00	39.00	1.60	3.60
X59T	2088	19.92	17.00	10.00	12.18	3	97	11.00	17.00	25.00	1.49	2.74
X61T	2088	23.11	19.00	14.00	12.42	7	94	15.00	19.00	28.00	1.79	3.79

[a]n=sample size
[b]SD=standard deviation

3.1 Descriptive Statistics

Prior to finding appropriate distributions, a thorough analysis of the descriptive statistics was performed on the daily concentrations of PM2.5 measured at the designated 12 sites in Bangkok. The findings of this examination are shown and condensed in Figure 1 and Table 2. Figure 1 presents the boxplots that depict the daily PM2.5 values measured at the specified station in Bangkok. The graphic depiction clearly shows that the distribution of daily PM2.5 values observed at all stations has a positive skewness. Table 2 reveals that station X54T had the highest average daily concentrations of PM2.5, with a maximum value of 32.56 $\mu g/m^3$. In contrast, station X59T registered the minimum average daily concentrations of PM2.5, measuring 19.92 $\mu g/m^3$. Furthermore, station X54T had the greatest median daily concentrations of PM2.5 at 29.00 $\mu g/m^3$, whereas stations X05T, X10T, and X59T had the lowest median daily concentrations of PM2.5 of 17.00 $\mu g/m^3$. Based on the remaining output presented in Table 2, it clearly shows that the distribution of daily PM2.5 values observed at all sites has a noticeable positive skewness.

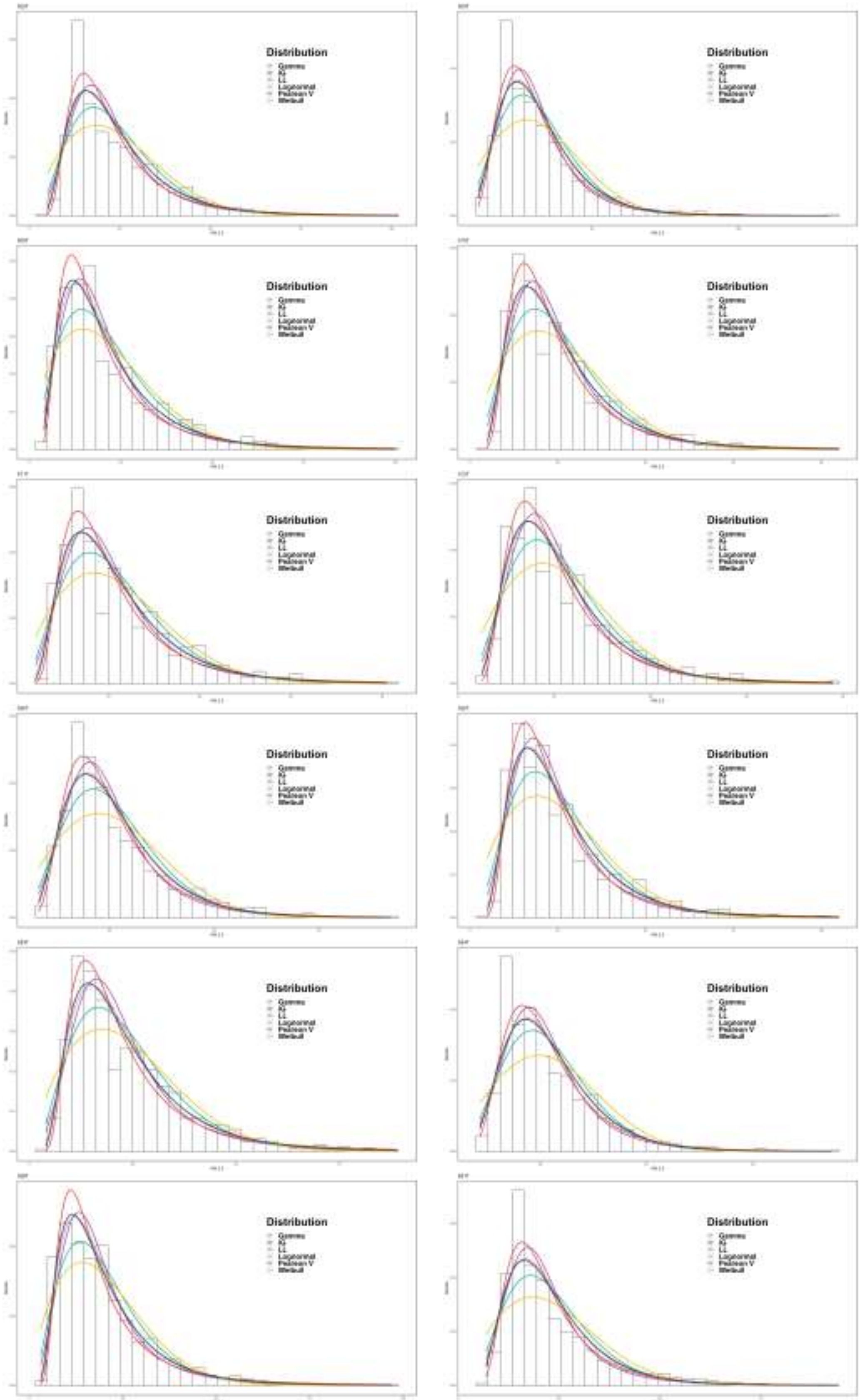

Fig. 2: Comparison of histograms and theoretical densities for statistical distributions evaluated on daily PM2.5 values observed at the Bangkok station

Table 3. The AD goodness-of-fit test result for six distribution functions of daily PM2.5 values in Bangkok

Station	Distribution					
	gamma	**IG**	**LN**	**LL**	**Weibull**	**Pearson V**
X02T	14.631	6.083	7.156	8.875	23.628	**4.213**
X03T	21.523	11.581	11.541	11.852	42.110	**6.258**
X05T	18.310	**3.440**	5.179	7.850	32.483	5.388
X10T	11.488	**3.165**	4.219	6.265	22.027	4.004
X11T	14.434	**5.918**	7.229	9.661	24.191	6.028
X12T	8.421	**2.152**	2.772	4.229	18.537	2.255
X50T	28.853	13.992	14.442	14.562	51.731	**6.665**
X52T	27.031	9.046	10.399	11.791	46.220	**4.032**
X53T	12.935	**2.425**	3.577	6.458	26.899	4.716
X54T	18.762	9.583	9.665	10.154	40.511	**4.830**
X59T	12.921	**2.353**	3.508	6.389	27.071	6.449
X61T	43.266	23.075	22.877	20.196	69.617	**11.095**

Table 4. Estimated parameters of the fitted distributions of daily PM2.5 for the indicated station in Bangkok

Station	Fitted Distribution	Estimated parameters	Station	Fitted Distribution	Estimated parameters
X02T	Pearson V	$\hat{\alpha} = 4.19782$, $\hat{\beta} = 77.40105$	X50T	Pearson V	$\hat{\alpha} = 5.94554$, $\hat{\beta} = 128.68040$
X03T	Pearson V	$\hat{\alpha} = 6.66742$, $\hat{\beta} = 178.79450$	X52T	Pearson V	$\hat{\alpha} = 4.03954$, $\hat{\beta} = 78.71426$
X05T	IG	$\hat{\mu} = 21.65233$, $\hat{\theta} = 50.79283$	X53T	IG	$\hat{\mu} = 23.57180$, $\hat{\theta} = 66.94004$
X10T	IG	$\hat{\mu} = 20.54397$, $\hat{\theta} = 62.31721$	X54T	Pearson V	$\hat{\alpha} = 7.66919$, $\hat{\beta} = 216.63070$
X11T	IG	$\hat{\mu} = 21.84124$, $\hat{\theta} = 69.83626$	X59T	IG	$\hat{\mu} = 19.92205$, $\hat{\theta} = 51.45492$
X12T	IG	$\hat{\mu} = 21.40408$, $\hat{\theta} = 81.00040$	X61T	Pearson V	$\hat{\alpha} = 5.37718$, $\hat{\beta} = 100.73330$

3.2 Distributions of Daily Concentrations of PM 2.5

Figure 2 displays histograms that show daily concentrations of PM2.5 measured at 12 sites. The histograms also include the fitted distributions for each of the six models that were investigated. The IG, LL, LN, and Pearson V distributions show a clear and high agreement with the data histogram of all stations. Table 3 presents the AD goodness-of-fit test statistic for six distributions. The smaller values in the table imply a better match with the actual daily concentrations of PM2.5 data. The daily PM2.5 values at the monitored stations, namely X02T, X03T, X50T, X52T, X54T, and X61T, showed a strong correlation with the Pearson V distribution. The IG distribution was determined to be the most suitable statistical distribution for the daily concentrations of PM2.5 at stations X05T, X10T, X11T, X12T, X53T, and X59T. The maximum likelihood technique is used to estimate the parameters that describe the distributions of daily PM2.5 values. This estimation is done using the "fitdist" function from the "fitdistrplus" package in the R language. The estimated parameters for the distributions of PM2.5 values fitted to each station are displayed in Table 4.

3.3 Percentiles and Exceedance Probabilities

The administrative targets for air pollution management typically range from the 98.0th to 99.9th percentile, [10], [11]. Therefore, it is crucial to carefully determine the distribution of daily PM2.5 levels in the higher range. Table 5 presents estimated values for the 98th percentile of PM2.5

levels, as well as the probabilities of detecting levels higher than 75 $\mu g/m^3$ at certain sites. The predicted 98th percentile of daily PM2.5 levels for stations X03T and X52T are 76.03 and 76.09 $\mu g/m^3$, respectively. These values exceed the 24-hour threshold for daily concentrations of PM2.5 in Thailand, which is set at 75 $\mu g/m^3$. The daily concentrations of PM2.5 at all stations have predicted 98th percentile values that exceed the USA EPA's 24-hour standard of 35 $\mu g/m^3$, [28]. Consequently, the threshold of 75 $\mu g/m^3$ (in Thailand) and 35 $\mu g/m^3$ (in the USA) was surpassed by the top two percent of the projected one-year data. This percentile is the equivalent of 7 days out of a year.

To analyze the Thailand 24-hour standard, we calculate the probability of finding levels higher than 75 $\mu g/m^3$, denoted as P(PM2.5>75), based on the Thailand air quality criteria provided in [5] as the average concentrations of PM2.5 during 24 hours. The probability of exceeding certain thresholds is displayed in Table 5. The data indicates that station X12T has the lowest probability, specifically 0.0019, while station X52T has the highest probability, specifically 0.0209.

Table 5. Estimated 98th percentile for PM2.5 levels (in $\mu g/m^3$) and the probabilities for detecting levels greater than 75 $\mu g/m^3$ at the indicated stations and using the indicated distribution

Station	**Fitted Distribution**	**PM2.5 (in $\mu g/m^3$)**		***P*(PM2.5>75)**
		Measured	**Predicted**	
X02T	Pearson V	60.00	67.81	0.0155
X03T	Pearson V	74.00	76.03	0.0166
X05T	IG	61.00	62.44	0.0083
X10T	IG	54.00	52.57	0.0030
X11T	IG	56.92	55.29	0.0038
X12T	IG	52.00	52.53	0.0019
X50T	Pearson V	61.00	65.85	0.0090
X52T	Pearson V	70.00	76.09	0.0209
X53T	IG	63.00	63.22	0.0082
X54T	Pearson V	71.76	73.29	0.0139
X59T	IG	55.00	55.22	0.0041
X61T	Pearson V	60.00	58.19	0.0069

Table 6. The 95% confidence intervals for the median of daily PM2.5 at the specified 12 stations in Bangkok

Station	**PB confidence intervals**	**SB confidence intervals**
X02T	(19.30765, 20.72887)	(19.27113, 20.69235)
X03T	(27.52335, 28.92723)	(27.07277, 28.47665)
X05T	(17.34427, 18.52679)	(15.47321, 16.65573)
X10T	(17.07891, 18.32170)	(15.67830, 16.92109)
X11T	(18.31270, 19.58816)	(16.41184, 17.68730)
X12T	(18.29317, 19.61972)	(18.38028, 19.70683)
X50T	(22.39272, 23.44950)	(20.55050, 21.60728)
X52T	(20.63952, 21.79127)	(20.20873, 21.36048)
X53T	(19.48881, 20.74965)	(19.25035, 20.51119)
X54T	(28.86097, 30.19331)	(27.80669, 29.13903)
X59T	(16.22465, 17.29611)	(16.70389, 17.77535)
X61T	(19.48698, 20.43835)	(17.56165, 18.51302)

3.4 Confidence Intervals for the Median of Daily PM2.5 Levels

The confidence intervals for the median of daily PM2.5 readings at the 12 designated stations in Bangkok are displayed in Table 6. The procedure for constructing parametric PB and SB confidence intervals for the median is outlined in the preceding section. The findings indicate that the majority of stations, except X03T and X54T, exhibit air quality that is deemed acceptable. However, it is important to note that certain individuals, particularly those who are very susceptible to air pollution, may still face potential health risks. Stations X03T and X54T have unhealthy air quality for sensitive groups. Individuals who are part of sensitive groups may have adverse health impacts, whereas the general public is less likely to be impacted.

4 Conclusion

Within this paper, an analysis was conducted on daily concentrations of PM2.5 values at twelve air quality monitoring sites in Bangkok. Six two-parameter statistical distribution models were used to identify the most appropriate distribution for accurately representing daily concentrations of PM2.5 data. The confidence intervals for the median of daily PM2.5 at each station have been computed. The daily concentrations of PM2.5 data at stations X02T, X03T, X50T, X52T, X54T, and X61T followed a Pearson V distribution. After careful analysis, it was concluded that the IG distribution is the most appropriate statistical distribution for stations X05T, X10T, X11T, X12T, X53T, and X59T. The PM2.5 data at all sites have been predicted to surpass the 98th percentile values of the USA EPA's 24-hour threshold of 35 $\mu g/m^3$. Moreover, the projected 98th percentile values above the 24-hour threshold of 75 $\mu g/m^3$ for daily concentrations of PM2.5 in Thailand at stations X03T and X52T. Confidence intervals indicate that stations X03T and X54T have air quality that is detrimental to sensitive populations and may have negative health effects based on the median of daily PM2.5 results.

This paper specifically focuses on the limitations of two-parameter statistical distributions. Future research should focus on analyzing distributions with more than two parameters and prioritize investigating the correlation between PM2.5 and meteorological factors. Furthermore, it is important to examine the annual mortality rates related to PM2.5 in significantly affected areas like Bangkok or other regions in Thailand that are experiencing significant PM2.5 issues.

Acknowledgement:
We extend our gratitude to the referees for their valuable suggestions on the manuscript and appreciate the helpful guidance provided by Associate Professor Patchanok Srisuradetchai. The authors gratefully acknowledge the financial support provided by the Faculty of Science and Technology, Thammasat University, Contract No. 1/2566.

References:

[1] A. Garcia, E. Santa-Helena, A. De Falco, J. de Paula Ribeiro, A. Gioda, and C. R. Gioda, "Toxicological Effects of Fine Particulate Matter (PM2.5): Health Risks and Associated Systemic Injuries—Systematic Review," *Water Air Soil Pollut.*, vol. 234, no. 6, p. 346, 2023, doi: https://doi.org/10.1007/s11270-023-06278-9.

[2] Health Effects Institute, "State of Global Air 2019," Boston, 2019, [Online]. https://www.stateofglobalair.org/sites/default/files/soga_2019_report.pdf (Accessed Date: July 13, 2023).

[3] N. R. Fold, M. R. Allison, B. C. Wood, P. T. B. Thao, S. Bonnet, S. Garivait, R. Kamens, and S. Pengjan, "An Assessment of Annual Mortality Attributable to Ambient PM2.5 in Bangkok, Thailand," *Int J Environ. Res. Public Health*, vol. 17, no. 19, p. 7298, Oct. 2020, doi: https://doi.org/10.3390/ijerph17197298.

[4] IQAir, "The 2022 World Air Quality Report: Region & City PM2.5 Ranking," 2022, [Online]. https://www.greenpeace.org/static/planet4-thailand-stateless/2023/03/d1d69c24-2022_world_air_quality_report_en.pdf (Accessed Date: July 13, 2023).

[5] P. Suraswadi, "Pollution Control Department announces Thailand's 2023 Air Quality Index," [Online]. https://ratchakitcha.soc.go.th/documents/140D157S0000000000300.pdf?fbclid=IwAR2bJGm_x9Df4ZBCOu-T9_RqDUdd27EhzGBL1jjhuU1lb-Wo6k6t4YKXsxM (Accessed Date: January 25, 2024).

[6] S. Wasinrat and B. Choopradit, "The Poisson Inverse Pareto Distribution and Its Application," *Thailand Statistician*, vol. 21,

no. 1, pp. 110–124, 2023, [Online]. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/248027 (Accessed Date: June 25, 2023).

[7] B. Choopradit and S. Wasinrat, "The Poisson-Paralogistic Distribution: Properties and Applications," *Journal of Applied Science and Emerging Technology*, vol. 22, no. 1, p. e249656, Apr. 2023, [Online]. https://ph01.tci-thaijo.org/index.php/JASCI/article/view/249656 (Accessed Date: June 25, 2023).

[8] W. E. Wilson, J. C. Chow, C. Claiborn, W. Fusheng, J. Engelbrecht, and J. G. Watson, "Monitoring of particulate matter outdoors," *Chemosphere*, vol. 49, no. 9, pp. 1009-1043, 2002, doi: https://doi.org/10.1016/S0045-6535(02)00270-9.

[9] H.-C. Lu and G.-C. Fang, "Estimating the frequency distributions of PM10 and PM2.5 by the statistics of wind speed at Sha-Lu, Taiwan," *Science of The Total Environment*, vol. 298, no. 1, pp. 119–130, 2002, doi: https://doi.org/10.1016/S0048-9697(02)00164-X.

[10] F. Karaca, O. Alagha, and F. Ertürk, "Statistical characterization of atmospheric PM10 and PM2.5 concentrations at a non-impacted suburban site of Istanbul, Turkey," *Chemosphere*, vol. 59, no. 8, pp. 1183–1190, 2005, doi: https://doi.org/10.1016/j.chemosphere.2004.11.062.

[11] C. Marchant, V. Leiva, M. F. Cavieres, and A. Sanhueza, "Air contaminant statistical distributions with application to PM10 in Santiago, Chile," in *Reviews of Environmental Contamination and Toxicology Volume 223*, in Reviews of Environmental Contamination and Toxicology. , Springer Science and Business Media, LLC, 2013, pp. 1–31. doi: https://doi.org/10.1007/978-1-4614-5577-6_1.

[12] W. Panichkitkosolkul and P. Srisuradetchai, "Bootstrap Confidence Intervals for the Parameter of Zero-truncated Poisson-Ishita Distribution," *Thailand Statistician*, vol. 20, no. 4, pp. 918–927, Sep. 2022, [Online]. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/247474 (Accessed Date: July 5, 2023).

[13] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "On the variety of methods for calculating confidence intervals by bootstrapping," *Journal of Animal Ecology*, vol. 84, no. 4, pp. 892–897, 2015, doi: https://doi.org/10.1111/1365-2656.12382.

[14] J. Carpenter and J. Bithell, "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians," *Stat Med*, vol. 19, no. 9, p. 1141—1164, May 2000, doi: https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.

[15] T. J. DiCiccio and B. Efron, "Bootstrap Confidence Intervals," *Statistical Science*, vol. 11, no. 3, pp. 189–212, 1996, [Online]. http://www.jstor.org/stable/2246110 (Accessed Date: July 23, 2023).

[16] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, 1st ed. Chapman and Hall/CRC, 1994.

[17] Pollution Control Department, "Daily PM2.5 concentration data for the 12 air quality monitoring stations in Bangkok," [Online]. http://air4thai.pcd.go.th/webV3/#/History (Accessed Date: May 31, 2023).

[18] S. Klugman, H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, 5th ed. Wiley, 2019.

[19] M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*, 3rd ed. New York: Wiley, 2000.

[20] G. E. Crooks, "The Amoroso Distribution," *arXiv: Statistics Theory*, 2010, [Online]. https://doi.org/10.48550/arXiv.1005.3274 (Accessed Date: June 09, 2023).

[21] R. Chhikara and L. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, 1st ed. CRC Press, 2019.

[22] J. Wang, L. Zhang, X. Niu, and Z. Liu, "Effects of PM2.5 on health and economic loss: Evidence from Beijing-Tianjin-Hebei region of China," *J Clean Prod*, vol. 257, p. 120605, 2020, doi: https://doi.org/10.1016/j.jclepro.2020.120605.

[23] M. L. Delignette-Muller and C. Dutang, "fitdistrplus: An R Package for Fitting Distributions," *J Stat Softw*, vol. 64, no. 4, pp. 1–34, 2015, doi: https://doi.org/10.18637/jss.v064.i04.

[24] M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," *J Am Stat Assoc*, vol. 69, no. 347, pp. 730–737, 1974, doi:

https://doi.org/10.1080/01621459.1974.10480196.

[25] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193 – 212, 1952, doi: https://doi.org/10.1214/aoms/1177729437.

[26] A. Agresti and B. Finlay, *Statistical Methods for the Social Sciences*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1997.

[27] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1 – 26, 1979, doi: https://doi.org/10.1214/aos/1176344552.

[28] U.S. Environmental Protection Agency, "Final Reconsideration of the National Ambient Air Quality Standards for Particulate Matter (PM)," [Online]. https://www.epa.gov/system/files/documents/2024-02/pm-naaqs-overview.pdf (Accessed Date: February 09, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Boonyarit Choopradit: Conceived the research, provided an original idea of the study, provided methodology, analyzed the data, programmed, provided a description, interpreted the data, and wrote the paper.
- Rujapa Paitoon, Nattawadee Srinuan, and Satita Kwankaew: Selected research data, and reviewed the paper.

All authors discussed the results and contributed to the final manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This article was funded by the Faculty of Science and Technology, Thammasat University, Contract No. 1/2566.

Conflict of Interest

The authors have no conflicts of interest to declare.