STATISTICAL MODELS OF $PM_{2.5}$ AND PM_{10} CONCENTRATIONS IN THE BANGKOK AREA

TANATIP HANPAYAK

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Mathematics

Suranaree University of Technology

Academic Year 2023

แบบจำลองทางสถิติของความเข้มข้น $PM_{2.5}$ และ PM_{10} บริเวณกรุงเทพมหานคร

นางสาวธนาทิพย์ หาญพยัคฆ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2566

STATISTICAL MODELS OF $PM_{2.5}$ AND PM_{10} CONCENTRATIONS IN THE BANGKOK AREA

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee

____

(Asst. Prof. Dr. Prapart Pue-on)

Chairperson

____

(Asst. Prof. Dr. Tidarut Areerak)

Member (Thesis Advisor)

____

(Assoc. Prof. Dr. Eckart Robert Schulz)

Member

____

(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)

Vice Rector for Academic Affairs

and Quality Assurance

____

(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

ธนาทิพย์ หาญพยัคฆ์ : แบบจำลองทางสถิติของความเข้มข้น $PM_{2.5}$ และ PM_{10} บริเวณกรุงเทพมหานคร (STATISTICAL MODELS OF $PM_{2.5}$ AND PM_{10} CONCENTRATIONS IN THE BANGKOK AREA) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.ธิดารัตน์ อารีรักษ์, 115 หน้า.

คำสำคัญ : ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน/ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน/การแจกแจงทางสถิติ/การแจกแจงผสม/การวิเคราะห์ค่าสุดขีด

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการแจกแจงที่เหมาะสมสำหรับอธิบายข้อมูลมลพิษทางอากาศอันเนื่องมาจากฝุ่นละออง $PM_{2.5}$ และ PM_{10} ในประเทศไทยโดยเฉพาะกรุงเทพมหานคร ทั้งนี้การวิเคราะห์ทางสถิติของข้อมูลมลพิษทางอากาศและการสร้างแบบจำลองผ่านการใช้โปรแกรมสำเร็จรูป RStudio โดยใช้ชุดข้อมูลจากฐานข้อมูลสาธารณะกรมควบคุมมลพิษ ได้แก่ ข้อมูลความเข้มข้นฝุ่นละออง $PM_{2.5}$ และ PM_{10} เฉลี่ย 24 ชั่วโมง ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2561 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 จำนวน 10 สถานีที่ตั้งอยู่ในพื้นที่บริเวณกรุงเทพมหานคร งานวิจัยนี้เริ่มจากการนำชุดข้อมูลมาศึกษาการแจกแจงทางสถิติพร้อมตรวจสอบความสอดคล้องของชุดข้อมูลกับการแจกแจงดังกล่าวโดยใช้การทดสอบภาวะสารูปดีและใช้เกณฑ์ข้อมูลในการค้นหาการแจกแจงที่ดีที่สุด พบว่าการแจกแจงล็อกนอร์มัลมีความเหมาะสมที่สุดในการสร้างแบบจำลอง จากนั้นนำชุดข้อมูลมาศึกษาการแจกแจงผสมโดยเลือกศึกษาการแจกแจงผสมสองแบบพบว่าการแจกแจงผสมสองแบบของล็อกนอร์มัลมีความเหมาะสมที่สุด และสำหรับการแจกแจงผสมสามแบบ พบว่าการแจกแจงผสมสามแบบของล็อกนอร์มัลและการแจกแจงแกมมา-แกมมา-ล็อกนอร์มัลมีความเหมาะสมที่สุด โดยใช้เกณฑ์ข้อมูล ท้ายที่สุดศึกษาการวิเคราะห์ค่าสุดขีดโดยเลือกศึกษาการแจกแจงสุดขีดวางนัยทั่วไปและการแจกแจงพาเรโตวางนัยทั่วไปเพื่อคาดการณ์ระดับการเกิดซ้ำของค่าสุดขีดของความเข้มข้นฝุ่นละออง $PM_{2.5}$ และ PM_{10} ในอีก 2 ปี 5 ปี 10 ปี และ 15 ปี ข้างหน้า โดยใช้การทดสอบภาวะสารูปดีและใช้เกณฑ์ข้อมูลในการตัดสินใจเลือกแบบจำลอง ทั้งนี้ซึ่งข้อสรุปจากการศึกษานี้สามารถใช้เป็นแนวทางในการวางแผนการจัดการและป้องกันปัญหาฝุ่นละออง $PM_{2.5}$ และ PM_{10} ในพื้นที่กรุงเทพมหานครรวมถึงบริเวณใกล้เคียงได้

สาขาวิชาคณิตศาสตร์และภูมิสารสนเทศ
ปีการศึกษา 2566

ลายมือชื่อนักศึกษา ธนาทิพย์ หาญพยัคฆ์
ลายมือชื่ออาจารย์ที่ปรึกษา ธิดารัตน์ อารีรักษ์

TANATIP HANPAYAK : STATISTICAL MODELS OF $PM_{2.5}$ AND PM_{10} CONCENTRATIONS IN THE BANGKOK AREA. THESIS ADVISOR : ASST. PROF. TIDARUT AREERAK, Ph.D. 115 PP.

Keyword: $PM_{2.5}$, PM_{10}, STATISTICAL DISTRIBUTIONS, MIXTURE DISTRIBUTIONS, EXTREME VALUE ANALYSIS

The objective of this research was to study appropriate distributions for describing air pollutant data, particularly $PM_{2.5}$ and PM_{10}, in Thailand, with focus on Bangkok. Statistical analysis of air pollution data and model building were conducted using the RStudio software. The dataset was obtained from the Pollution Control Department's public database, comprised of 24-hour average $PM_{2.5}$ and PM_{10} concentration data from January 1, 2018, to December 31, 2022, and collected from 10 stations located in the Bangkok metropolitan area. The research began by examining non-mixture distributions and assessing the goodness-of-fit of the dataset to these distributions and the information criteria. The best fit distribution was found to be the log-normal distribution. Subsequently, the dataset was analyzed using mixture distributions, including 2-mixture distributions and 3-mixture distributions. It was determined that the 2-mixture lognormal distribution, the 3-mixture lognormal distribution and the gamma-gamma-lognormal distribution were the most suitable based on the information criteria. Finally, extreme value analysis was conducted to predict the maximum pollution levels, including $PM_{2.5}$ and PM_{10}, for the next 2, 5, 10, and 15 years. This analysis involved studying the generalized extreme value distribution and the generalized Pareto distribution to estimate return levels. Model selection was based on goodness-of-fit and the information criteria. The conclusions drawn from this study can serve as guidelines for planning management and prevention strategies for $PM_{2.5}$ and PM_{10} pollution issues in Bangkok and its surrounding areas.

School of Mathematical Sciences and Geoinformatics

Academic Year 2023

Student's Signature Tanatip

Advisor's Signature T. Areerak

ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to all who have given me advice, and whose unwavering guidance and support illuminated the path to the completion of this thesis. Foremost among them is Asst. Prof. Dr. Tidarut Areerak, my thesis advisor, whose invaluable assistance transcended mere academic direction. From refining presentation skills to crafting a compelling paper, Prof. Areerak consistently lent her expertise and attentiveness to my evolving ideas, for which I am profoundly thankful.

My supporters have been equally instrumental in this academic journey. Asst. Prof. Dr. Jessada Tanthanuch provided astute counsel on structuring the thesis, which was indispensable. Without collaborative efforts, this thesis would not have reached fruition. For this invaluable assistance, I am deeply indebted.

A sincere appreciation extends to the member of School of Mathematical Sciences and Geoinformatics at the Institute of Science, Suranaree University of Technology (SUT). Their commitment to imparting profound knowledge has been instrumental in shaping my academic pursuits, and I am grateful for the enriching learning experience.

Lastly, I express my profound thanks for being a recipient of the graduate studies scholarship for outstanding academic performers from Suranaree University of Technology, which facilitated my pursuit of a master’s degree. This financial support played a pivotal role in realizing my academic aspirations. In conclusion, I extend my deepest gratitude to all who contributed to the successful completion of this thesis. Your support has been the cornerstone of my achievement, and for that, I am truly thankful.

Tanatip Hanpayak

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

Page

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

Page

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

Page

LIST OF TABLES (Continued)

Page

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Air pollution has a negative impact on human health and life expectancy, while also exerting manifold, far-reaching effects on society and the economy. The harmful effects of air pollution not only endanger human health but also have a wide range of negative economic effects, such as rising healthcare expenses and decreased worker productivity, making it a complicated problem with significant societal and economic implications (Erickson et al., 2016).

Particulate matter (PM) is defined by WHO as the most significant air pollutant, with $PM_{2.5}$ and PM_{10} representing particulate matter, where particles have an aerodynamic diameter equal to or less than 2.5 μm and equal to or less than 10 μm, respectively. $PM_{2.5}$ and PM_{10} are composed mainly of varying amounts of water and several major components, including sulfates, acids, nitrates, elemental carbon, organic carbon, and trace metals, depending on their sources (Almeida et al., 2006). In South and Southeast Asia, from 1999 to 2014, a multitude of issues arose. During this timeframe, the region witnessed a substantial 38% increase in premature mortality attributed to PM, with the total number of multi-year premature deaths in South-Southeast Asia due to $PM_{2.5}$ reaching 1,447,000. The primary health issues associated with $PM_{2.5}$ exposure were stroke and ischemic heart disease. Notably, India and Bangladesh emerged as the primary contributors to the mortality burden caused by $PM_{2.5}$ in South and Southeast Asia. It is worth noting that South Asia recorded a higher estimate of premature deaths compared to Southeast Asia during the 1999-2014 period (Shi et al., 2018). In Thailand, the Pollution Control Department (PCD) found that overall air quality has shown improvement compared to the previous year. However, certain areas still experienced levels of $PM_{2.5}$ and PM_{10} dust that exceeded permissible limits. In 2022, the five provinces facing the most significant air quality challenges were Saraburi, Samut Songkhram, Bangkok, Phitsanulok, and Nong Khai, with the number of days surpassing air quality standards being 97, 77, 68, 57, and

54, respectively.

In this thesis, we will focus on the concentrations of $PM_{2.5}$ and PM_{10} in Bangkok. Significant sources of $PM_{2.5}$ and PM_{10} pollution in Bangkok have been identified. Concerning PM_{10}, it is first of all attributed to re-suspended soil and cooking, constituting 10% to 15%, with notable contributions from automotive emissions and biomass burning, accounting for roughly 22% and 28%, respectively. In the case of $PM_{2.5}$, the primary culprits are automobiles (32%), biomass burning (26%), meat preparation (31%), and road dust (6%) (Chuersuwan et al., 2008).

Taylor, Jakeman and Simpson (1986) investigated the distribution of air pollution particles, specifically ozone, carbon monoxide, sulfur dioxide, oxides of nitrogen, nitrogen dioxide, and nitrogen. These particles are major components of $PM_{2.5}$ and PM_{10}. They observed that these air pollution particles follow lognormal, gamma, and Weibull distributions. Consequently, several researchers have used these distributions, as well as mixture distributions, to analyze the distributions of $PM_{2.5}$ and PM_{10}. For example, in a study by Xi et al. (2013), it was found that the lognormal distribution provided the best fit for the daily PM_{10} concentration distributions in Beijing, Guangzhou, Shanghai, Wuhan, and Xi'an between 2004 and 2008, as assessed through goodness-of-fit tests, including chi-square, Kolmogorov-Smirnov (KS), and Anderson-Darling (AD) tests. In another study by Chu, Yu, and Kuo (2012), a finite mixture distribution model (FMDM) was identified as the most suitable mixture distribution for monthly $PM_{2.5}$ and PM_{10} data from a dust storm that occurred in Taiwan in March 2008. The study of statistical distribution models, including the mixture distribution model, enables us to effectively capture the diversity present in the data. By combining several statistical distribution models, we can address the complexity of the various sources contributing to PM2.5 and PM10 pollutants. Specifically, our focus is on Bangkok, where numerous locations contribute to these pollutants. These include ozone, carbon monoxide, sulfur dioxide, oxides of nitrogen, nitrogen dioxide, and nitrogen, which are major components of PM2.5 and PM10. Since these pollutants originate from various sources, the mixture distribution model helps identify these diverse pollution sources.

Nevertheless, understanding the pollution distribution alone does not help the public or the government to fully address the issue. Being aware of the trends in maximum pollution levels can help raise public awareness of pollution issues. Therefore, we have chosen to investigate extreme value analysis, which can estimate the return level of $PM_{2.5}$ and PM_{10}. For instance, Martins et al. (2017) proposed applying extreme value analysis to hourly pollutant data measured in Sao Paulo (MASP) over sixteen years (1996 to 2011) and hourly pollutant data measured in Rio de Janeiro (MARJ) over seven years (2005 to 2011). They found that MASP had a higher probability of extreme events compared to MARJ, indicating a shorter return period.

The goal of the thesis is to analyze the appropriate statistical distributions and mixture distributions for $PM_{2.5}$ and PM_{10} concentrations in Bangkok. The maximum likelihood method is used for parameter estimation. Goodness-of-fit is utilized to analyze $PM_{2.5}$ and PM_{10} concentration data, fitting them to various statistical distributions and also making comparisons among these distributions, Additionally, we aim to apply extreme value analysis to $PM_{2.5}$ and PM_{10} data to determine the return level.

1.1 Research Objective

The objectives of the research are as follows:

1. To find the statistical distribution for $PM_{2.5}$ and PM_{10} concentration data.
2. To apply extreme value analysis for analyzing $PM_{2.5}$ and PM_{10} concentration data.

1.2 Scope and Limitations

The scopes of the research work are as follows:

1. The data on daily average $PM_{2.5}$ and PM_{10} concentrations in Bangkok was obtained from the Pollution Control Department, Air Quality and Noise Management Bureau, The Ministry of Natural Resources and Environment of Thailand, as of 22 July 2021.
2. The R programming language was used in the research.

1.3 Research Procedure

The research work proceeded as follows:

1. Study statistical distributions, including Lognormal distribution, Gamma distribution, and Weibull distribution, as well as mixture distributions.

2. Study parameter estimation and goodness-of-fit.

3. Use goodness-of-fit to analyze $PM_{2.5}$ and PM_{10} concentration data fitted to some statistical distributions and also make comparisons among different distributions.

4. Study and apply extreme value theory analysis for $PM_{2.5}$ and PM_{10} concentration data.

1.4 Results Obtained

The results of the research work are as follows:

1. The appropriate statistical distribution for $PM_{2.5}$ and PM_{10} concentrations in Bangkok has been identified.

2. The return level and return period of $PM_{2.5}$ and PM_{10} concentrations in Bangkok has been analyzed.

CHAPTER II
LITERATURE REVIEW

This chapter gives an overview of basic mathematical ideas, including statistics.The main references for this information are Kvam and Vidakovic (2007), Krishnamoorthy (2006) and Coles (2001).

2.1 Statistical Distributions

Well-known distributions and distributions commonly employed in this thesis are presented as follows.

2.1.1 Gamma Distribution

A gamma random variable X with shape parameter a and scale parameter b is denoted by

$$X \sim \text{gamma}(a, b).$$

For any $x > 0$, the probability density function of the gamma distribution is given by

$$f(x; a, b) = \frac{1}{\Gamma(a) b^a} e^{-\frac{x}{b}} x^{a-1}, \tag{2.1}$$

where parameters a and b are positive real numbers and $\Gamma(a)$ is the gamma function. The cumulative distribution function corresponding to (2.1) is

$$F(x; a, b) = \frac{1}{\Gamma(a)} \int_0^{x/b} e^{-t} t^{a-1} dt. \tag{2.2}$$

Consider the events generated by a Poisson distribution with parameter λ. The waiting time for the first event to occur is represented by an exponential random variable and the waiting time for the n^{th} event to occur is represented by the gamma random variables X. Therefore,

$$X = \sum_{i=1}^{n} Y_i,$$

where $Y_1, ..., Y_n$ are independent exponential random variables with a parameter $1/\lambda$. The gamma random variable can be seen as an extension of the exponential random variable.

The gamma distribution is commonly employed to describe situations in which one is concerned with the waiting time for a finite number of independent events to occur. This assumption is based on events happening at a constant rate, and the probability of more than one event occurring in a short period is exceedingly low. Application areas for this distribution include dependability and queuing theory. Examples include the distribution of component failure times, the distribution of calibration intervals for equipment that requires recalibration after a set number of uses, and the distribution of customer wait times for a given value of k. Additionally, the gamma distribution can be used to predict daily rainfall totals in a region (Krishnamoorthy, 2006).

2.1.2 Lognormal Distribution

The lognormal positive random variable X with parameters μ and σ is denoted as

$$X \sim \text{lognormal}(\mu, \sigma^2),$$

when $Y = \ln(X)$, the Y is a normal random variable with a mean μ and a standard deviation σ. For any $x > 0$, the probability density function of the lognormal distribution is given by

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \tag{2.3}$$

where the parameter σ is a positive real number, and $-\infty < \mu < \infty$. The cumulative distribution function corresponding to (2.3) is

$$\begin{aligned} F(x; \mu, \sigma) &= P(X \leq x \mid \mu, \sigma) \\ &= P(\ln X \leq \ln x \mid \mu, \sigma) \\ &= P\left(Z \leq \frac{\ln x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \end{aligned} \tag{2.4}$$

where Φ represents the standard normal distribution function, and we have defined $z = (\ln x - \mu)/\sigma$.

When the random variable X takes only positive values and exhibits a significantly right-skewed histogram, the lognormal distribution might be suggested in physical contexts. In particular, the natural logarithmic transformation of the data must satisfy the normality assumption for the lognormal model to be relevant for a physical problem. Nevertheless, in many real-world scenarios, the lognormal and gamma distributions can be used interchangeably. The lognormal distribution can be used for modeling raindrop sizes, global position data, and wind speed (Krishnamoorthy, 2006).

2.1.3 Weibull Distribution

The Weibull random variable X with scale parameter σ and shape parameter η is denoted as

$$X \sim \text{Weibull}(\sigma, \eta).$$

For any $x > 0$, the probability density function of the Weibull distribution is given by

$$f(x; \sigma, \eta) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} e^{-\left(\frac{x}{\sigma}\right)^{\eta}}, \tag{2.5}$$

where parameter σ and η are positive real numbers. The cumulative distribution function corresponding to (2.5) is

$$F(x; \sigma, \eta) = 1 - e^{-\left(\frac{x}{\sigma}\right)^{\eta}}. \tag{2.6}$$

The Weibull distributions have found applications in a wide range of scientific disciplines, as seen by the numerous articles that demonstrate this. In systems engineering, the Weibull distribution is frequently used to examine the overall performance degradation of complex systems. It can generally be used to describe information about the length of time between events. This is how it is used in engineering, actuarial science, and risk analysis. Additionally, the Weibull distribution is also useful in the fields of biology, earth science, and medicine (Krishnamoorthy, 2006).

2.1.4 Mixture Distributions

When a population consists of heterogeneous subgroups, each represented by a separate probability distribution, mixture distributions arise. If the observer cannot distinguish between the sub-distributions based on the observation, they are left with an unsorted mixture. For example, a finite mixture of k distributions has the probability density function given by

$$f_X(x) = \sum_{i=1}^{k} p_i f_i(x), \tag{2.7}$$

where f_i represents the probability density function of subpopulation i, and the weights p_i represent the likelihood of an observation being generated from subpopulation i based on its probability density function. These weights, for $i = 1, ..., k$, satisfy the condition that $\sum_{i=1}^{k} p_i = 1$ (Kvam and Vidakovic, 2007).

2.2 Extreme Value Analysis

Extreme value analysis is a statistical approach used to study and model the behavior of extreme or rare events. It focuses on understanding the distribution of extreme values in a dataset. It is widely used in the context of environmental, financial, economics, and engineering data where extreme events, such as severe storms, financial market crashes, or structural failures, are of particular interest. Depending on the selection of data used in extreme values analysis, the distribution of extreme value theory can be categorized into two approaches, which are as follows:

1. **Block maxima model**

   The block maxima model is used to analyze data from a specific time period of interest, such as annually, monthly, weekly or daily. It involves selecting the data with the highest or lowest values within each time period for analysis in accordance with the generalized extreme value distribution (GEV).

2. **Peak over threshold model**

   In this model, data points that exceed a certain threshold are considered and analyzed. An essential step in the model is choosing a threshold that is appropriate for the data being analyzed. This threshold should take into account the non-independence of the data and may be adjusted by declustering values that exceed the threshold. This approach is suitable for the generalized Pareto distribution (GPD). The peak over threshold model is commonly used when dealing with large datasets or data collected daily.

2.2.1 Generalized Extreme Value Distribution

The block maximum model doesn't seem to have been widely used in statistical applications until the 1950s. The methodology was popularized by Gumbel in 1958. The parameterization of the generalized extreme value distribution for extreme value limit models was independently proposed by von Mises in 1954 and Jenkinson in 1955 (Coles, 2001).

Consider $X_1, X_2, ..., X_n$ as independent random variables with a common distribution function F. The maximum value of the random variable is denoted as

$$M_n = \max\{X_1, ..., X_n\}.$$

The cumulative distribution function of the generalized extreme value distribution is given by

$$F(x;\mu,\sigma,\xi) = exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}, \tag{2.8}$$

where σ, ξ and μ represent the scale parameter, the shape parameter, and the location parameter, respectively. In this parameterization, for $\xi > 0$ and $\xi < 0$, the extreme value distributions are represented by the Fréchet distribution and the Weibull distribution, respectively. The Gumbel distribution, with its distribution function, is derived from a subset of the generalized extreme value distribution family where $\xi = 0$ and can be seen as the limit of equation (2.8) as $\xi \to 0$.

2.2.2 Generalized Pareto Distribution

The generalized Pareto distribution is often used to model the distribution of extreme values that exceed a certain threshold, which is a common approach because it is often impractical to directly model the extreme tail of a distribution using standard methods. The generalized Pareto distribution is especially useful for modeling the tails of distributions because it is flexible and can accurately describe the distribution of extreme values (Coles, 2001).

The use of the generalized Pareto distribution for modeling excesses over a high threshold is justified by arguments on the asymptotic behavior of the data. Let us consider $X_1, X_2, ..., X_n$ to be independent random variables with common distribution function F. For any $x > 0$, the cumulative distribution function of exceedances over the threshold u is defined by

$$P\left(X - u \leq x | X > u\right),$$

for $Y_i = X_i - u$, where $X_i > u$, the cumulative distribution function of $Y_1, \ldots, Y_{n_u}$ can be approximated by the generalized Pareto distribution given by

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma_\mu}\right)^{-1/\xi}, \tag{2.9}$$

for some ξ and positive real numbers σ and μ. This equation is defined on $y > 0$ and $(1 + \xi y/\sigma_\mu) > 0$, where

$$\sigma_\mu = \sigma + \xi(u - \mu).$$

2.2.3 Return Level and Return Period

Return levels and return periods are fundamental concepts in extreme value analysis, aiding in the assessment and management of extreme events. The return period or recurrence interval, T, is an estimate of the interval of time between events of a certain intensity or size (Rakhecha and Singh, 2009). The T year return level, x_T, is the level exceeded on average only once in every T years (Coles, 2001).

For $\xi \neq 0$, the return level of the generalized extreme value distribution is given by

$$x_T = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-\ln\left(1 - \frac{1}{T}\right)\right\}^{-\xi}\right],$$

where $F(x_T) = 1 - 1/T$ with σ, ξ and μ represent the scale parameter, the shape parameter, and the location parameter, respectively.

For $\xi \neq 0$, the return level of the generalized Pareto distribution is given by

$$x_m = \mu + \frac{\sigma}{\xi}\left[1 - \left(\frac{nT}{M}\right)^{-\xi}\right],$$

where n is the total number of excesses over the threshold u in M years with σ, ξ and μ represent the scale parameter, the shape parameter, and the location parameter, respectively (Abild et al., 1992).

2.3 Maximum Likelihood Estimation

In statistics, a variety of techniques are employed to estimate parameters, such as the method of moments, the probability-weighted moments method, and the maximum likelihood method. Mage and Ott (1984) investigated the efficiency of these three distinct techniques for parameter estimation using lognormal distributions of air pollutant concentrations. The maximum likelihood estimation consistently provides the most accurate parameter estimates. Therefore, for the purpose of parameter estimation in this research, the maximum likelihood method will be chosen.

Let X be a random variable with a probability density function $f(x;\theta)$ dependent on an unknown parameter θ from a sample space Ω. Consider a random sample $X_1, X_2, \ldots, X_n$ from X, where $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with a common probability density function $f(x;\theta)$. The likelihood function, denoted as $L(\theta;x)$ for $x = (x_1, x_2, \ldots, x_n)$, is defined as the product of the individual probability density values:

$$L(\theta;x) = \prod_{i=1}^{n} f(x_i;\theta),\ \theta \in \Omega. \tag{2.10}$$

This function is commonly represented as $L(\theta)$. The maximum likelihood estimator (MLE) of θ, denoted as $\hat{\theta}$, is the value that maximizes the likelihood function $L(\theta)$. However, it is usually more convenient to work with the natural logarithm of the likelihood function, known as the log-likelihood function:

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta),\ \theta \in \Omega. \tag{2.11}$$

The maximum likelihood estimator can be obtained by solving the equation:

$$\frac{\partial l(\theta)}{\partial \theta} = 0. \tag{2.12}$$

In cases where this equation is nonlinear, a numerical approach can be applied to find the solution for equation (2.12) (Hogg et al., 2019).

2.4 Tools for Evaluating Models

2.4.1 Goodness-of-Fit

In this section, we present techniques for testing the hypothesized distribution to fit the data, known as goodness-of-fit. Pearson introduced this term in 1902 to describe statistical evaluations that assess how well a model or distribution matches a given set of data. The first method for assessing goodness-of-fit for general distributions was developed by Kolmogorov in 1933.

In goodness-of-fit testing, two hypotheses related to an unknown parameter in the underlying distribution of the data are considered. One is known as the null hypothesis, denoted as H_0, which represents the hypothesis that the observed data follows a specific hypothesized distribution. The other is the alternative hypothesis, denoted as H_1 or H_a, representing the hypothesis that the observed data does not conform to the hypothesized distribution. Errors can arise when making conclusions based on these hypotheses. There are two types of errors as follows (Kvam and Vidakovic, 2007).

Type I error: Type I error occurs when the null hypothesis H_0 is incorrectly rejected when it is actually true. The probability of this type of error is denoted by α and refers to the significance level of the test.

Type II error: Type II error occurs when the null hypothesis H_0 is accepted when it is actually false. The probability of this type is represented by β, and $1-\beta$ defines the power of the test. In simple terms, the power of a test measures its ability to correctly reject incorrect alternative hypotheses.

In the thesis, various tests for assessing goodness-of-fit are considered, including the Kolmogorov-Smirnov Test, the Anderson-Darling Test, and the Cramér-Von Mises Test.

Kolmogorov-Smirnov Test

Andrei Nikolaevich Kolmogorov stands as one of the most accomplished and famous mathematicians in history. His significant contributions to probability theory include the development of test statistics for distribution functions, some of which are now named after him. Another mathematician, Nikolai Vasil'yevich Smirnov, extended Kolmogorov's work by introducing the Smirnov test for comparing two samples. This section focuses on the Kolmogorov-Smirnov test, which is used to evaluate the suitability of distribution functions for fitting the data. The Kolmogorov-Smirnov test serves as the foundation for many nonparametric goodness-of-fit tests for distributions (Kvam and Vidakovic, 2007).

Consider a dataset $X_1, X_2, \ldots, X_n$ sampled from a population with an unknown cumulative distribution function, denoted as F. In the context of hypothesis testing, we have the null hypothesis:

$$H_0 : F(x) = F_0(x), \forall x,$$

versus the alternative hypothesis:

$$H_1 : F(x) \neq F_0(x),$$

where F_0 represents the expected cumulative distribution function. The statistic for the Kolmogorov-Smirnov test is denoted as D_n and is defined as:

$$D_n = \sup_t |F_n(t) - F_0(t)|, \tag{2.13}$$

where F_n is the empirical cumulative distribution function based on the sample. Furthermore, the modified statistic is presented as:

$$\sqrt{n}D_n = \sup_x \sqrt{n} |F_n(x) - F_0(x)| .$$

Anderson-Darling Test

In an effort to enhance the Kolmogorov-Smirnov statistic, Anderson and Darling modified the statistic for specific distributions in 1954, which is known as the Anderson-Darling statistic. The Anderson-Darling test is employed to determine whether a sample of data came from a population with a specific distribution. The critical values of the

Anderson-Darling test are determined based on the specified distribution. This refinement enhances the test, but the disadvantage is that critical values must be determined for each hypothesized distribution. In contrast, the Kolmogorov-Smirnov test is distribution free, meaning its critical values are independent of the specific distribution being tested. Let us consider a dataset $X_1, X_2, \ldots, X_n$ collected from a population with an unknown cumulative distribution function F. For hypothesis testing, we have the null hypothesis:

$$H_0 : F(x) = F_0(x), \forall x,$$

versus the alternative hypothesis:

$$H_1 : F(x) \neq F_0(x),$$

where F_0 represents the expected cumulative distribution function. The statistic for the Anderson-Darling test is denoted as A^2 and is defined as follows:

$$A^2 = -n - S, \tag{2.14}$$

where

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[F_0(X_{i:n}) + \ln\left(1 - F_0\left(X_{n+1-i:n}\right)\right)\right],$$

and $X_{i:n}$ is the ordered sample values, which $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ (Kvam and Vidakovic, 2007).

Cramér-Von Mises Test

Harald Cramér and Richard von Mises proposed the statistic for a goodness-of-fit test known as the Cramér-Von Mises statistic. This statistic measures the weighted distance between the empirical cumulative distribution function and the expected cumulative distribution function. It is based on a squared-error function. Let us consider a dataset $X_1, X_2, \ldots, X_n$ obtained from a population with an unknown cumulative distribution function F. For the purpose of hypothesis testing, we have the null hypothesis:

$$H_0 : F(x) = F_0(x), \forall x,$$

versus the alternative hypothesis:

$$H_1 : F(x) \neq F_0(x),$$

where F_0 represents the expected cumulative distribution function. The Cramér-Von Mises statistic is defined as follows:

$$w_n{}^2(\psi(F_0)) = \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 \psi(F_0(x)) dF_0(x), \tag{2.15}$$

where F_n is the empirical cumulative distribution function based on the sample. There are several commonly used choices for the weight functional ψ. In the case of $\psi(x) = 1$, this is called the "standard" Cramér-Von Mises statistic

$$w_n{}^2(1) = w_n{}^2,$$

in which case the test statistic becomes

$$n w_n{}^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F_0(X_{i:n}) - \frac{2i-1}{2n} \right)^2. \tag{2.16}$$

In the case of $\psi(x) = x^{-1}(1-x)^{-1}$, the test statistic becomes

$$w_n{}^2(1/(F_0(1-F_0))) = A^2/n,$$

where A^2 represents the Anderson-Darling statistic (Kvam and Vidakovic, 2007).

2.4.2 Information Criterion

Choosing the optimal model for a given dataset involves striking the appropriate balance between model fit and complexity. This issue is tackled through the application of two widely acknowledged criteria:

1. Akaike's Information Criterion (AIC),
2. Bayesian Information Criterion (BIC).

Following are the definitions for these criteria.

Akaike Information Criterion

The Akaike Information Criterion (AIC) is a model selection principle proposed by Akaike. It aims to estimate out-of-sample prediction loss by combining the in-sample prediction loss with a correction term. This concept was elucidated by Ding, Tarokh, and Yang in 2018. The AIC value of the model is defined by

$$\text{AIC} := -2l + 2k$$

where l is the maximized value of the loglikelihood for the model with k representing the number of estimated parameters in the model.

Bayesian Information Criterion

The Bayesian Information Criterion (BIC) is another popular model selection principle. The key distinction from AIC is that the penalty term, instead of a constant 2, is replaced with the logarithm of the sample size (Ding, Tarokh, and Yang, 2018). The BIC value of the model is defined by

$$\text{BIC} := -2l + (\ln(n))k$$

where n is the sample size. Therefore, as is commonly practiced, one calculates the AIC and BIC for each model and selects the model with the smallest criterion value.

2.5 Related Researches

Lu (2004) conducted a study on modeling PM_{10} data in central Taiwan. Three different distributions, namely lognormal, Weibull, and gamma distributions, were used to model the data. The distribution parameters were estimated using the maximum likelihood method. Mean absolute error, root mean square error, and the index of agreement were employed to determine the appropriateness of the distributions. The results indicated that the gamma distribution is the most suitable for representing high PM_{10} concentrations. However, it sometimes diverges when predicting high PM_{10} concentrations accurately. To address this issue, two predicting methods were introduced: the two-parameter exponential distribution and the asymptotic distribution of extreme values. Both predicting methods effectively estimate return periods and exceedances over a critical concentration in the future.

Mijić et al. (2009) analyzed PM_{10} concentrations in the air measured in the Belgrade urban area between 2003 and 2005. Lognormal, Weibull, and type V Pearson distributions were employed for this purpose. The method of least squares and the method of moments were used to calculate the parameters of the distributions. However, they found a divergence in forecasting a high PM_{10} concentration. Therefore, extreme value distributions were chosen to fit the high PM_{10} concentration distribution. This approach enables the prediction of the return period and exceedances over critical concentration levels in subsequent years.

Chu, Yu and Kuo (2012) utilized a finite mixture distribution model (FMDM) to analyze monthly $PM_{2.5}$ and PM_{10} data from a dust storm that occurred in Taiwan in March 2008. Sequential Gaussian simulation (SGS) was employed with FMDM cut-off values for $PM_{2.5}$ and PM_{10} to map the probabilities of a contaminated area. The results indicate that both $PM_{2.5}$ and PM_{10} can be accurately modeled using the FMDM, and SGS with FMDM cut-off values helps in identifying areas with both high and low concentrations.

Xi et al. (2013) studied the statistical distribution characteristics of daily average PM_{10} concentration in Beijing, Guangzhou, Shanghai, Wuhan, and Xi'an between 2004-2008. Lognormal, Weibull, and gamma distributions were used to identify the PM_{10} concentration distribution. The maximum likelihood approach was applied to find distribution parameters. The goodness-of-fit tests such as chi-square, Kolmogorov-Smirnov (KS), and Anderson-Darling (AD) were employed to determine whether the distribution is appropriate. The results demonstrate that when the three distributions were compared, the lognormal distribution was found to be the best daily PM_{10} concentration distribution in the 5 cities, as assessed by three goodness-of-fit tests.

Martins et al. (2017) suggested employing extreme value analysis on hourly pollutant data measured in Sao Paulo (MASP) over sixteen years (1996 to 2011) and on hourly pollutant data measured in Rio de Janeiro (MARJ) over seven years (2005 to 2011). This analysis included determining the probabilities of exceedance and the return period for high concentrations of pollutant, including carbon monoxide (CO), nitrogen oxides (NO and NO_2), O_3, and PM, in both regions. The study utilized two approaches: firstly, the generalized extreme value distribution (GEV), which was applied to the monthly maximum hourly concentration measurements. Secondly, the generalized Pareto distribution (GPD), commonly used to model the tails of distribution composed of values exceeding a threshold, was used. Specifically, GPD was used for daily maxima above a threshold, while GEV was used for monthly maxima data. The results indicated that, despite GEV and GPD being different approaches, they presented similar results.

Plocoste et al. (2020) investigated the modeling of PM_{10} frequency distribution and extreme events in the Caribbean basin. The analysis involved testing five distinct distributions: lognormal, Weibull, Burr, stable, and mixture distributions. The Kolmogorov-Smirnov test was used to assess the appropriateness of the distribution by categorizing the study into the low dust season, high dust season, and extreme events. In summary, they found that the Burr and Weibull mixture model was appropriate for modeling both classical and extreme events.

Mishra et al. (2021) investigated $PM_{2.5}$ distributions in five countries: India, China, France, Brazil, and the United States of America (USA). They compared $PM_{2.5}$ concentrations during their respective lockdown period in 2020 and the corresponding period in 2019. Their hypothesis is that the lockdown significantly reduced city traffic and human activity, likely resulting in decreased pollution levels. This study helps us understand the limitations and necessary steps to improve air quality. For the analysis of $PM_{2.5}$ in these five countries, lognormal, Weibull, and gamma distributions were employed. The maximum likelihood method was used to calculate the parameters of these distributions, and the chi-square test was used to assess their appropriateness. In summary, when compared to all other probability functions used in this study, the gamma distribution provides the best fit for the $PM_{2.5}$ concentration, according to the chi-square test. Furthermore, it was observed that the mean value of $PM_{2.5}$ concentration decreased during the 2020 lockdown period compared to the same period of 2019.

CHAPTER III
RESEARCH METHODOLOGY

This chapter presents an overview of the research methodology used to identify statistical models for $PM_{2.5}$ and PM_{10} data. The study examined the statistical distributions of daily average $PM_{2.5}$ and PM_{10} concentrations in Bangkok. The following sections are included:

1. Preparing data,
2. Employing distribution and mixture distributions for data fitting,
3. Employing extreme value analysis for data fitting.

3.1 Preparing Data

This thesis uses daily average $PM_{2.5}$ and PM_{10} concentrations from 10 stations in Bangkok, as listed in Table 3.1 and illustrated in Figure 3.1, covering the period from 2018 to 2022. These data were collected by the Pollution Control Department, Air Quality and Noise Management Bureau, Ministry of Natural Resources and Environment of Thailand. The dataset, available on https://pcd.gdcatalog.go.th/ and last updated in July 2021, provides detailed information on pollutant concentrations, including $PM_{2.5}$ and PM_{10} concentrations, with variables outlined in Table 3.2 and example data shown in Table 3.3.

The research utilizes RStudio version 4.3.2, running on the Microsoft Windows 11 operating system (version 22H2).

3.2 Employing Distribution and Mixture Distributions for Data Fitting

To select data for analysis, both types of distributions that is non-mixture and mixture distributions consider data from daily average $PM_{2.5}$ and PM_{10} concentrations divided into stations and years.

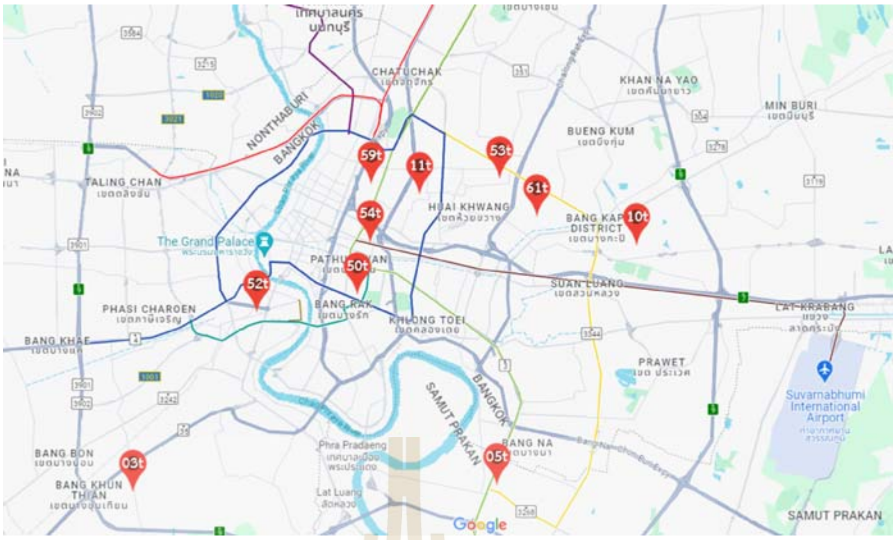

Figure 3.1 Location of air quality monitoring stations in Bangkok.

3.2.1 Non-Mixture Distribution

The analysis of statistical distributions involves considering three distributions: gamma distribution (GM), lognormal distribution (LN), and Weibull distribution (W) are considered. The parameters of each distribution are determined using the maximum likelihood method. Subsequently, the suitability of the data for the selected distribution is evaluated through the goodness-of-fit tests, including the Kolmogorov-Smirnov test (KS), the Anderson-Darling test (AD), and the Cramér-Von Mises test (CM). These tests provide statistic values and p-values. Additionally, information criteria, namely Akaike information criteria (AIC) and Bayesian information criteria (BIC), are utilized in the evaluation of distributions. R programming, along with the “fitdistrplus” and “stats” packages is employed in this study to estimate parameters and evaluate information criteria, test the goodness-of-fit.

3.2.2 Mixture Distributions

For the analysis of mixture distributions, gamma distribution, lognormal distribution, and Weibull distribution are combined into 2-mixture distributions and 3-mixture distributions, as illustrated in Table 3.4. The parameters of each mixture distribution

Table 3.1 List of chosen stations for air quality monitoring in Bangkok used in the thesis.

Code	Station name
05t	Thai Meteorological Department
10t	Khlong Chan Community Housing
11t	Huai Khwang Community Housing Authority
59t	Public Relations Department
61t	Bodindecha School
03t	Along Highway No. 3902
50t	Chulalongkorn Hospital
52t	Thonburi Sub-Electricity Authority
53t	Chokchai Metropolitan Police Station
54t	Din Daeng Community Housing

Table 3.2 Variables of particulate matter data used in the thesis.

Variable	Description
$PM_{2.5}$	daily average $PM_{2.5}$ concentrations ($\mu g/m^3$)
PM_{10}	daily average PM_{10} concentrations ($\mu g/m^3$)

are determined using the maximum likelihood method. Following this, the appropriateness of the data for the selected distribution is evaluated using the information criteria, specifically Akaike information criteria (AIC) and Bayesian information criteria (BIC). R programming with "ltmix" packages is employed in this study to estimate parameters and evaluate information criteria.

3.3 Employing Extreme Value Analysis for Data Fitting

For extreme value analysis, two distributions are considered: generalized extreme value distribution (GEV) and generalized Pareto distribution (GPD). The utilization of data for these two distributions varies slightly, leading to the division of data selection as follows

Table 3.3 Example of PM_{10} concentration data.

Date	Concentration ($\mu g/m^3$)									
	05t	10t	11t	59t	61t	03t	50t	52t	53t	54t
31/07/2020	28	20	29	22	27	55	36	27	29	56
01/08/2020	27	19	27	22	25	47	36	27	24	61
02/08/2020	21	17	25	22	25	46	33	24	24	58
03/08/2020	17	12	26	18	20	38	28	18	19	55
04/08/2020	13	13	32	18	23	60	31	21	21	49
05/08/2020	20	15	31	17	26	72	38	26	23	60
06/08/2020	31	18	31	19	26	90	37	34	27	53
07/08/2020	25	16	30	21	24	74	34	27	23	60
08/08/2020	31	18	29	23	29	84	43	32	26	59
09/08/2020	15	13	24	18	22	47	25	17	17	49

3.3.1 Generalized Extreme Value Distribution (GEV)

To select data for analysis, GEV considers data from a specified time period of interest. We aggregated data on the concentrations of $PM_{2.5}$ and PM_{10} from 2018 to 2022. Subsequently, we then segmented the data into monthly intervals, by selecting the highest value of each month for analysis.

3.3.2 Generalized Pareto Distribution (GPD)

To select data for analysis, GPD considers data that exceeds a certain threshold. The selection of threshold typically focuses on the largest values in the dataset to extract relevant information. Quantiles are utilized to partition the data into intervals, with particular emphasis often placed on the highest quantile. As a result, we examined three thresholds derived from the quantiles: 90%, 95%, and 99%. For analysis, we selected the values that exceeded these threshold.

The parameters of the distribution are determined using the maximum likelihood

Table 3.4 Mixture Distributions.

Type	Distribution
2-Mixture Distributions	2-Gamma distributions (2GM)
	2-Lognormal distributions (2LN)
	2-Weibull distributions (2W)
	Gamma Lognormal distribution (GM-LN)
	Gamma Weibull distribution (GM-W)
	Lognormal Weibull distribution (LN-W)
3-Mixture Distributions	3-Gamma distributions (3GM)
	3-Lognormal distributions (3LN)
	3-Weibull distributions (3W)
	Gamma Gamma Lognormal distribution (GM-GM-LN)
	Gamma Gamma Weibull distribution (GM-GM-W)
	Gamma Lognormal Lognormal distribution (GM-LN-LN)
	Gamma Lognormal Weibull distribution (GM-LN-W)
	Gamma Weibull Weibull distribution (GM-W-W)
	Lognormal Lognormal Weibull distribution (LN-LN-W)
	Lognormal Weibull Weibull distribution (LN-W-W)

method. Then, the goodness-of-fit tests, Anderson-Darling test (AD) and Cramér-Von Mises test (CM), which produce statistic values and p-values, and the information criteria, Akaike information criteria (AIC) and Bayesian information criteria (BIC), are employed to determine whether the data are suitable for the selected distribution. The purpose of this is determine the return level. R programming is used to study extreme value analysis with several packages, including “stats”, “extRemes”, and “gnFit”. These packages are utilized to examine threshold, estimate parameters, information criteria and return levels, and evaluate goodness-of-fit tests, respectively.

This chapter has presented an overview of the research methodology used to identify statistical dtributions for $PM_{2.5}$ and PM_{10} data. The study examines the statistical distribution of daily average $PM_{2.5}$ and PM_{10} concentrations in Bangkok. A summary of the methods are clearly presented in Figure 3.2.

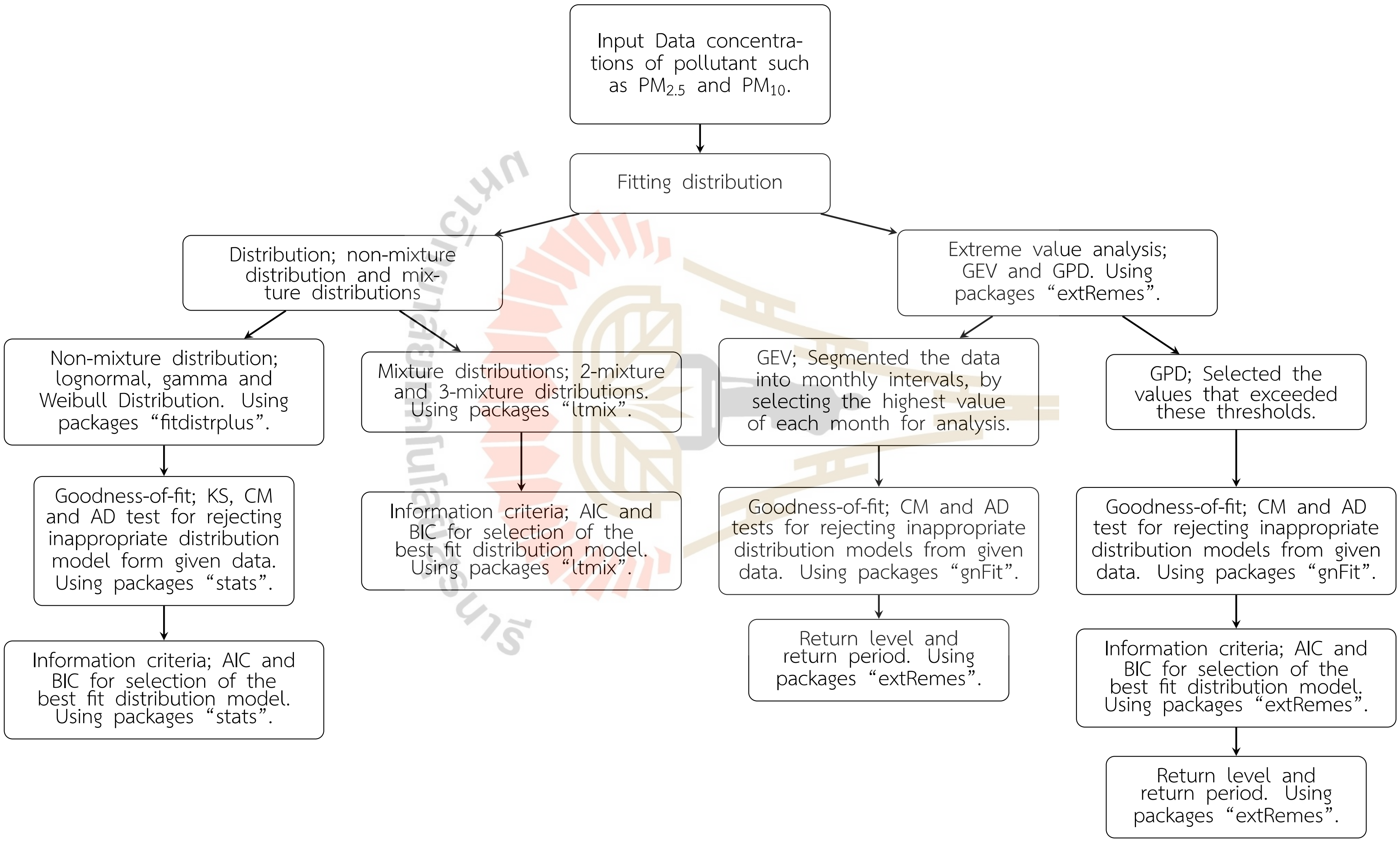

Figure 3.2 The summary of all methodology.

CHAPTER IV
RESULTS AND DISCUSSION

The first part of our analysis focuses on the concentrations of $PM_{2.5}$ and PM_{10} data sets categorized by years, specifically from 2018 to 2022, spanning the period from January 1, 2018, to December 31, 2022, for the ten previously mentioned stations in Bangkok. In this study, distribution functions were used to characterize the statistical distribution of $PM_{2.5}$ and PM_{10}, which have been determined to follow gamma, lognormal, and Weibull distributions. The distributions of $PM_{2.5}$ and PM_{10} concentrations were examined in order to obtain the best representative distributions by using goodness-of-fit and information criteria.

A significance level of 0.05 was used for the goodness-of-fit test. If the p-value is lower than 0.05, the null hypothesis will be "rejected". Conversely, if the p-value is greater than 0.05, the null hypothesis will be "non-rejected". Regarding information criteria, the model with the lowest information criteria score indicates a better fit compared to other models.

In this thesis, for the goodness-of-fit test, a distribution is defined as "non-rejected" if at least one test result in non-rejection, and "rejected" if all tests result in rejection or if the distribution cannot be estimated. For information criteria, a "1" in the table indicates the best model supported by the information criteria.

4.1 Result of Non-Mixture Distribution

The analysis of gamma, lognormal, and Weibull distribution models using R software includes that the information criteria included Akaike information criteria (AIC) and Bayesian information criteria (BIC), while goodness-of-fit tests comprised the Kolmogorov-Smirnov (KS), Cramér-Von Mises (CM), and Anderson-Darling (AD) tests. These serve to select the best distribution and evaluate the appropriateness of distributions, respec-

Table 4.1 Results of the goodness-of-fit tests (KS, CM, AD) of non-mixture distributions for $PM_{2.5}$ and PM_{10} from 2018 to 2022.

		$PM_{2.5}$		PM_{10}	
Year	Distribution	Non-rejected	Rejected	Non-rejected	Rejected
2018	Gamma	4	6	3	7
	Lognormal	7	3	7	3
	Weibull	4	6	2	8
2019	Gamma	2	8	6	4
	Lognormal	10	0	10	0
	Weibull	0	10	1	9
2020	Gamma	0	10	0	10
	Lognormal	3	7	1	9
	Weibull	0	10	0	10
2021	Gamma	1	9	1	9
	Lognormal	3	7	2	8
	Weibull	0	10	0	10
2022	Gamma	2	8	6	4
	Lognormal	4	6	7	3
	Weibull	0	10	3	7

tively. The statistics and p-values for goodness-of-fit, as well as the information criteria, are presented sample result in Table A.1 to A.15 for $PM_{2.5}$ and Table A.16 to A.30 for PM_{10} in the Appendix.

4.1.1 Goodness-of-fit of Non-Mixture Distribution

For goodness-of-fit tests of $PM_{2.5}$ and PM_{10} from 2018 to 2022, the comprehensive results are provided in Table 4.1. The numbers in the table indicate the number of stations in each year that “rejected” or “non-rejected” the considered distribution. Additionally, the summary results for the entire periods from 2018 to 2022 are present in Table 4.2. Based on the result of the goodness-of-fit tests for $PM_{2.5}$ and PM_{10}, the lognormal distribution performed the best, followed by the gamma distribution and the Weibull distribution, respectively.

Table 4.2 Summary results of the goodness-of-fit tests (KS, CM, AD) of non-mixture distributions for $PM_{2.5}$ and PM_{10} for the entire period.

	$PM_{2.5}$		PM_{10}	
Distribution	Non-rejected	Rejected	Non-rejected	Rejected
Gamma	9	41	16	34
Lognormal	27	23	27	23
Weibull	4	46	6	44

Table 4.3 Results of the information criteria (AIC and BIC) of non-mixture distribution for $PM_{2.5}$ and PM_{10} from 2018 to 2022.

	$PM_{2.5}$						PM_{10}					
	Gamma		Lognormal		Weibull		Gamma		Lognormal		Weibull	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	1	1	9	9	0	0	1	1	9	9	0	0
2019	0	0	10	10	0	0	0	0	10	10	0	0
2020	0	0	10	10	0	0	0	0	10	10	0	0
2021	0	0	10	10	0	0	0	0	10	10	0	0
2022	0	0	10	10	0	0	1	1	8	8	1	1

4.1.2 Information Criteria of Non-Mixture Distribution

The result of the information criteria (AIC and BIC) for $PM_{2.5}$ and PM_{10} from 2018 to 2022, are presented in Table 4.3. The numbers in the table represent the number of stations in each year that best fit the considered distribution (that is the lowest value of the information criteria). Additional, the summary results of the entire period 2018 to 2022 are shown in Table 4.4. Consequently, for both the AIC and BIC of $PM_{2.5}$ and PM_{10}, the lognormal distribution, is identified as the best fit followed by the gamma distribution and lastly, the Weibull distribution.

Table 4.4 Summary results of the information criteria (AIC and BIC) of non-mixture distribution for $PM_{2.5}$ and PM_{10} for the entire period.

	$PM_{2.5}$		PM_{10}	
Distribution	AIC	BIC	AIC	BIC
Gamma	1	1	2	2
Lognormal	49	49	47	47
Weibull	0	0	1	1

4.2 Result of Mixture Distributions

For analysis of the mixture distribution models using R software, the information criteria (AIC and BIC) were used to select the best model. The analysis of the mixture distributions is separated into two parts: the 2-mixture distributions and the 3-mixture distributions. Sample result of information criteria is provided in Table **??** to **??** for the 2-mixture distributions and Table A.41 to A.42 for the 3-mixture distributions in the Appendix.

4.2.1 Information Criteria of the 2-Mixture Distributions

The analysis of the 2-mixture distribution models included the 2-mixture gamma distribution (2GM), 2-mixture lognormal distribution (2LN), 2-mixture Weibull distribution (2W), lognormal-gamma distribution (LN-GM), lognormal-Weibull distribution (LN-W), and gamma-Weibull distribution (GM-W). This analysis utilized the information criteria such as AIC and BIC to determine the best distribution. Additionally, the sample results of the information criteria are provided in Table A.31 to A.40 in the Appendix.

The results of the information criteria (AIC and BIC) of $PM_{2.5}$ and PM_{10} are shown in Table 4.5 and 4.6, respectively. Table 4.7 summarizes thee results of AIC and BIC for the 2-mixture distribution of $PM_{2.5}$ and PM_{10} for entire period of 2018 to 2022. Accordingly, the 2-mixture lognormal distribution is identified as the best fit, followed by the lognormal-gamma distribution, and finally, the 2-mixture gamma distribution.

Table 4.5 Results of the information criteria (AIC and BIC) of the 2-mixture distribution for $PM_{2.5}$ from 2018 to 2022.

	2GM		2LN		2W		LN-GM		LN-W		LN-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	2	2	3	3	0	0	5	5	0	0	0	0
2019	1	1	4	4	0	0	5	5	0	0	0	0
2020	3	3	2	2	0	0	5	5	0	0	0	0
2021	0	0	9	9	0	0	1	1	0	0	0	0
2022	0	0	8	8	0	0	2	2	0	0	0	0

Table 4.6 Results of the information criteria (AIC and BIC) of the 2-mixture distribution for PM_{10} from 2018 to 2022.

	2GM		2LN		2W		LN-GM		LN-W		LN-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	1	1	6	6	0	0	3	3	0	0	0	0
2019	0	0	8	8	0	0	2	2	0	0	0	0
2020	0	0	3	3	0	0	7	7	0	0	0	0
2021	1	1	9	9	0	0	0	0	0	0	0	0
2022	0	0	8	8	0	0	2	2	0	0	0	0

Table 4.7 Summary results of the information criteria (AIC and BIC) of 2-mixture distribution for $PM_{2.5}$ and PM_{10} for the entire period in 2018 to 2022.

	$PM_{2.5}$		PM_{10}	
Distribution	AIC	BIC	AIC	BIC
2GM	6	6	2	2
2LN	26	26	34	34
2W	0	0	0	0
LN-GM	18	18	14	14
LN-W	0	0	0	0
GM-W	0	0	0	0

Table 4.8 Results of the information criteria (AIC and BIC) of the 3-mixture distribution for $PM_{2.5}$ from 2018 to 2022.

Distribution;	3GM		3LN		3W		GM-GM-LN		LN-LN-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	4	4	1	1	0	0	5	4	0	0
2019	5	4	2	3	0	0	2	2	0	0
2020	2	3	4	4	0	0	2	1	0	0
2021	0	0	4	5	0	0	0	1	0	0
2022	0	0	5	5	0	0	1	1	0	0
Distribution;	GM-GM-W		GM-LN-LN		GM-LN-W		GM-W-W		LN-W-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	0	1	0	0	0	0	0	0	0	0
2019	0	0	1	1	0	0	0	0	0	0
2020	0	0	2	2	0	0	0	0	0	0
2021	0	0	6	4	0	0	0	0	0	0
2022	0	0	4	4	0	0	0	0	0	0

4.2.2 Information Criteria of the 3-Mixture Distributions

The analysis encompassed the 3-mixture distribution models: the 3-mixture gamma distribution (3GM), 3-mixture lognormal distribution (3LN), and 3-mixture Weibull distribution (3W). Additionally, it included hybrid models such as the gamma-gamma-lognormal distribution (GM-GM-LN), the gamma-gamma-Weibull distribution (GM-GM-W), the gamma-lognormal-lognormal distribution (GM-LN-LN), the gamma-lognormal-Weibull distribution (GM-LN-W), the gamma-Weibull-Weibull distribution (GM-W-W), the lognormal-lognormal-Weibull distribution (LN-LN-W), and the lognormal-Weibull-Weibull distribution (LN-W-W). This analysis utilized information criteria such as AIC and BIC to select the best distribution. Additionally, the sample results of information criteria is provided in Table A.41 to A.60 in the Appendix.

Table 4.9 Results of the information criteria (AIC and BIC) of the 3-mixture distribution for PM_{10} from 2018 to 2022.

Distribution;	3GM		3LN		3W		GM-GM-LN		LN-LN-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	4	4	4	4	0	0	1	1	0	0
2019	1	1	3	3	0	0	4	4	0	0
2020	1	1	4	4	0	0	4	4	0	0
2021	1	0	4	1	0	0	4	7	0	0
2022	1	2	2	1	0	0	5	4	0	0
Distribution;	GM-GM-W		GM-LN-LN		GM-LN-W		GM-W-W		LN-W-W	
Year	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2018	0	0	1	1	0	0	0	0	0	0
2019	0	0	2	2	0	0	0	0	0	0
2020	0	0	1	1	0	0	0	0	0	0
2021	0	0	1	2	0	0	0	0	0	0
2022	0	0	2	3	0	0	0	0	0	0

The result of the information criteria (AIC and BIC) for $PM_{2.5}$ and PM_{10} from 2018 to 2022 are indicated in Table 4.8 and 4.9, respectively. The summary results of the entire period 2018 to 2022 are shown in Table 4.10. In the case of $PM_{2.5}$, the 3-mixture lognormal distribution, followed by the gamma-lognormal-lognormal distribution, the 3-mixture gamma distribution, the gamma-gamma-lognormal distribution and the gamma-gamma-Weibull distribution, respectively.

In the case of PM_{10}, the gamma-gamma-lognormal distribution is the best fit, followed by the 3-mixture lognormal distribution, the 3-mixture gamma distribution, the gamma-lognormal-Weibull distribution, the gamma-gamma-Weibull distribution and the lognormal-lognormal-Weibull distribution, respectively.

Table 4.10 Summary of the results of the information criteria (AIC and BIC) of the 3-mixture distribution for $PM_{2.5}$ and PM_{10} for the entire period.

	$PM_{2.5}$		PM_{10}	
Distribution	AIC	BIC	AIC	BIC
3GM	11	11	8	8
3LN	16	18	17	13
3W	0	0	0	0
GM-GM-LN	10	9	18	20
GM-GM-W	0	1	0	0
GM-LN-LN	13	11	7	9
GM-LN-W	0	0	0	0
GM-W-W	0	0	0	0
LN-LN-W	0	0	0	0
LN-W-W	0	0	0	0

4.3 Result of Extreme Value Analysis

The extreme value analysis focuses on two distributions: the generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD). Information criteria, including the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), are utilized to select the best distribution and evaluate its appropriateness. Additionally, goodness-of-fit tests, such as the Cramér-Von Mises (CM), and the Anderson-Darling (AD) tests, are employed for the same purpose. The results of the generalized extreme value distribution and the generalized Pareto distribution are presented in Table 4.11 to 4.12 and 4.15 to 4.18, respectively.

4.3.1 The Generalized Extreme Value Distribution

For the analysis of the generalized extreme value distribution (GEV), the concentration data of PM2.5 and PM10 from 2018 to 2022 are segmented into monthly intervals, and the highest value of each month is selected for analysis.

For goodness-of-fit tests of $PM_{2.5}$ and PM_{10}, statistics and p-values for each station are provided in Table 4.11. Non-rejections at the 5% significance level are marked in bold.

Table 4.11 The outcomes of goodness-of-fit tests for GEV. The form shows the statistic (p-value) for $PM_{2.5}$ and PM_{10}. Non-rejections at the 5% level at above marked in bold.

Station code	GEV of $PM_{2.5}$		GEV of PM_{10}	
	CM (p-value)	AD (p-value)	CM (p-value)	AD (p-value)
05t	0.0632 **(0.3434)**	0.4086 **(0.3461)**	0.0537 **(0.4581)**	0.3246 **(0.5237)**
10t	0.0723 **(0.2610)**	0.4288 **(0.3101)**	0.0891 **(0.1581)**	0.5496 **(0.1571)**
11t	0.0728 **(0.2571)**	0.5122 **(0.1948)**	0.0255 **(0.9031)**	0.1699 **(0.9335)**
59t	0.0575 **(0.4082)**	0.4047 **(0.3534)**	0.0467 **(0.5608)**	0.353 **(0.4655)**
61t	0.099 **(0.1161)**	0.6258 **(0.1031)**	0.1705 (0.0129)	0.9164 (0.0198)
03t	0.045 **(0.5901)**	0.3214 **(0.5299)**	0.0192 **(0.97496)**	0.1475 **(0.9661)**
50t	0.1342 (0.0387)	0.7889 (0.0408)	0.0929 **(0.14095)**	0.5497 **(0.1570)**
52t	0.0805 **(0.2042)**	0.5127 **(0.1942)**	0.0662 **(0.3137)**	0.4376 **(0.2955)**
53t	0.0915 **(0.1472)**	0.6576 **(0.0861)**	0.0537 **(0.4581)**	0.349 **(0.47533)**
54t	0.0471 **(0.5542)**	0.3148 **(0.5442)**	0.0187 **(0.97845)**	0.1493 **(0.96395)**

Additionally, summary results are presented in Table 4.19. Based on of the goodness-of-fit tests, the generalized extreme value distribution appears appropriate for the data. Results of the information criteria are shown in Table 4.12. It is unnecessary to compare these criteria with other distributions as they suffice to determine if this distribution matches the data.

Table 4.12 The outcomes of the information criteria (AIC and BIC) for GEV of $PM_{2.5}$ and PM10.

Station code	GEV of $PM_{2.5}$		GEV of PM_{10}	
	AIC	BIC	AIC	BIC
05t	531.5961	537.8791	570.5732	576.8562
10t	430.3759	436.1714	561.5705	567.8032
11t	435.759	441.5545	338.2104	342.961
59t	511.5077	517.7907	543.0086	549.2917
61t	509.1075	515.3905	488.7223	494.5761
03t	465.4656	471.2611	525.844	531.6977
50t	520.8249	527.108	533.2505	539.3265
52t	536.6933	542.9763	580.7504	587.0335
53t	528.2222	534.5052	576.6669	582.95
54t	479.7894	485.7564	517.6278	523.5948

Table 4.13 The results of return level of GEV for $PM_{2.5}$.

$PM_{2.5}$	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	92.37324	118.06265	140.20047	154.38514
10t	74.88445	91.95498	105.95345	114.62620
11t	77.4001	94.0419	107.4092	115.5754
59t	93.03345	132.77080	172.80819	201.32604
61t	97.50345	142.87955	190.43152	225.21446
03t	107.5848	130.2768	148.5918	159.8173
50t	95.9062	124.2902	149.9845	167.0065
52t	111.6108	154.6651	196.6201	225.8363
53t	89.75225	110.57054	127.67775	138.29144
54t	101.9140	124.3456	142.9626	154.5910

Finally, the return levels of the generalized extreme value distribution of $PM_{2.5}$ and PM_{10} are shown in Table 4.13 - 4.14, considering 2-year period, 5-year period, 10-year period, and 15-year period.

Table 4.14 The results of return level of GEV for PM_{10}.

PM_{10}	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	130.0161	156.8206	178.2632	191.3266
10t	112.4959	125.5810	134.2214	138.8423
11t	122.0958	149.5646	172.2129	186.2964
59t	114.4871	144.9645	171.7479	189.1386
61t	120.6307	137.4411	149.5401	156.4009
03t	179.0756	196.9515	208.7989	215.1509
50t	140.7377	163.4765	180.8922	131.1934
52t	152.4232	191.5155	225.2905	246.9694
53t	142.4049	169.0427	190.0339	202.6924
54t	151.9136	168.1932	179.4976	185.7535

4.3.2 The Generalized Pareto Distribution

For the analysis of the generalized Pareto distribution (GPD) the concentration data of $PM_{2.5}$ and PM_{10} is utilized, spanning from 2018 to 2022. Threshold selection and quantiles play crucial roles in analyzing distributed data, especially those exhibiting heavy tail distributions commonly observed in datasets with extreme values. Threshold selection focuses on the largest values in the dataset to extract relevant information. Quantiles are utilized to partition the data into intervals, with particular emphasis often placed on the highest quantile, representing the most significant segment of the data or the "tail" of the distribution. In our study, three thresholds obtained from the quantiles: 90%, 95%, and 99%. The value that exceeds over the threshold are selected for analysis.

For goodness-of-fit tests, statistics and p-values for each station are provided in Table 4.15 for $PM_{2.5}$ and Table 4.16 for PM_{10}. Non-rejections at the 5% significance level are marked in bold. Additionally, summary results are presented in Table 4.19. Consequently, the GPD with a threshold obtained from quantile 95% has the highest number of non-rejections, followed by the GPD with a threshold obtained from quantile 99%, and finally, the GPD with a threshold obtained from quantile 90%.

Table 4.15 The outcomes of goodness-of-fit tests for GPD. The form shows the statistic (p-value) for $PM_{2.5}$. Non-rejections at the 5% level at above marked in bold.

Station code	GPD 90%		GPD 95%		GPD 99%	
	CM (p-value)	AD (p-value)	CM (p-value)	AD (p-value)	CM (p-value)	AD (p-value)
05t	0.1564 (0.01963)	1.1964 (0.00405)	0.1175 **(0.06488)**	0.8156 (0.0351)	0.041 **(0.6634)**	0.344 **(0.4879)**
10t	0.1474 (0.0258)	1.0696 (0.0083)	0.1051 **(0.09574)**	0.6277 **(0.10203)**	0.0322 **(0.81598)**	0.2983 **(0.58745)**
11t	0.1488 (0.02472)	1.165 (0.0048)	0.0636 **(0.3393)**	0.4854 **(0.2266)**	0.0788 **(0.2148)**	0.5652 **(0.1435)**
59t	0.1044 **(0.0979)**	0.8657 (0.0264)	0.0642 **(0.3332)**	0.4907 **(0.2199)**	0.0373 **(0.7313)**	0.2822 **(0.6373)**
61t	0.0585 **(0.3959)**	0.5465 **(0.1599)**	0.064 **(0.33519)**	0.481 **(0.2323)**	0.0401 **(0.6801)**	0.3045 **(0.5701)**
03t	0.0577 **(0.4057)**	0.4851 **(0.227)**	0.0382 **(0.7151)**	0.4039 **(0.3549)**	0.0834 **(0.1873)**	0.4936 **(0.2164)**
50t	0.1767 (0.0107)	1.3136 (0.0021)	0.0989 **(0.11648)**	0.6081 **(0.1141)**	0.0338 **(0.7910)**	0.2942 **(0.5996)**
52t	0.2596 (0.0009)	1.5605 (0.0005)	0.0384 **(0.71143)**	0.3375 **(0.5042)**	0.0428 **(0.6299)**	0.2652 **(0.6939)**
53t	0.2011 (0.0052)	1.3399 (0.0018)	0.0689 **(0.2891)**	0.4558 **(0.2673)**	0.0588 **(0.3924)**	0.3805 **(0.4024)**
54t	0.1197 **(0.0606)**	0.9544 **(0.01595)**	0.0533 **(0.46376)**	0.388 **(0.3866)**	0.069 **(0.2883)**	0.4379 **(0.2950)**

Results of the information criteria for $PM_{2.5}$ and PM_{10} are shown in Table 4.17 and Table 4.18, respectively. Summary results are presented in Table 4.20. Consequently, the GPD with a threshold obtained from quantile 99% has the the lowest values of information criteria, followed by the GPD with a threshold obtained from quantile 95%, and finally, the GPD with a threshold obtained from quantile 90%. The result obtained from analysis with information criteria align with those from the goodness-of-fit tests.

Table 4.16 The outcomes of goodness-of-fit tests for GPD. The form shows the statistic (p-value) for PM_{10}. Non-rejections at the 5% level at above marked in bold.

Station code	GPD 90%		GPD 95%		GPD 99%	
	CM (p-value)	AD (p-value)	CM (p-value)	AD (p-value)	CM (p-value)	AD (p-value)
05t	0.0778	0.5789	0.0684	0.5155	0.1499	0.8337
	(0.2213)	**(0.1324)**	**(0.2935)**	**(0.1911)**	(0.02391)	(0.03165)
10t	0.1017	0.7155	0.0379	0.4285	0.0322	0.2983
	(0.1066)	**(0.0619)**	**(0.7205)**	**(0.3106)**	**(0.8801)**	**(0.8277)**
11t	0.1147	0.8095	0.0407	0.2824	0.0488	0.3253
	(0.0708)	(0.0363)	**(0.66897)**	**(0.6367)**	**(0.5275)**	**(0.5225)**
59t	0.0747	0.5457	0.0608	0.5233	0.0317	0.249
	(0.2428)	(0.1607)	**(0.3693)**	**(0.1828)**	**(0.8234)**	**(0.74792)**
61t	0.0585	0.4885	0.0434	0.4321	0.0704	0.4359
	(0.39595)	**(0.2227)**	**(0.6189)**	**(0.3046)**	**(0.27636)**	**(0.2983)**
03t	0.0768	0.5777	0.0656	0.4778	0.0566	0.4013
	(0.2281)	**(0.1333)**	**(0.3194)**	**(0.2365)**	**(0.4195)**	**(0.3599)**
50t	0.0318	0.3192	0.0372	0.3143	0.0624	0.442
	(0.8219)	**(0.5344)**	**(0.7331)**	**(0.5453)**	**(0.3518)**	**(0.2885)**
52t	0.0641	0.4373	0.0571	0.4582	0.0395	0.2895
	(0.3342)	**(0.296)**	**(0.4131)**	**(0.2638)**	**(0.6912)**	**(0.6140)**
53t	0.0492	0.3988	0.0511	0.3388	0.0456	0.3925
	(0.5216)	**(0.3649)**	**(0.49593)**	**(0.5028)**	**(0.5796)**	**(0.3774)**
54t	0.0624	0.4777	0.0581	0.4403	0.1198	0.9285
	(0.3518)	**(0.2366)**	**(0.4008)**	**(0.2912)**	**(0.0604)**	(0.0185)

Finally, the return levels of the generalized Pareto distribution of $PM_{2.5}$ and PM_{10} are shown in Table 4.21 - 4.26, considering 2-year period, 5-year period, 10-year period, and 15-year period.

Table 4.17 The outcomes of the information criteria (AIC and BIC) for GPD of $PM_{2.5}$. The lowest values are marked in bold.

Station code	GPD 90%		GPD 95%		GPD 99%	
	AIC	BIC	AIC	BIC	AIC	BIC
05t	1151.76	1157.984	615.1089	620.0636	**112.3661**	**113.9113**
10t	955.74	961.6516	466.6116	471.1649	**84.62143**	**58.59954**
11t	999.124	1005.145	**92.24943**	**92.66553**	490.0154	494.6235
59t	1200.97	1207.277	573.834	578.6956	**109.5296**	**111.0748**
61t	1219.687	1225.947	595.3733	600.2586	**119.4333**	**121.214**
03t	1262.911	1268.985	**135.3924**	**135.9375**	615.0833	619.7448
50t	1203.182	1209.511	617.6537	622.631	**115.9814**	**117.8703**
52t	1298.802	1305.166	673.1953	678.217	**129.6531**	**131.542**
53t	1201.258	1207.553	603.6946	608.6719	**105.1697**	**106.9505**
54t	1143.919	1150.082	585.7472	590.5113	**120.0553**	**121.6005**

Table 4.18 The outcomes of the information criteria (AIC and BIC) for GPD of PM_{10}. The lowest values are marked in bold.

Station code	GPD 90%		GPD 95%		GPD 99%	
	AIC	BIC	AIC	BIC	AIC	BIC
05t	1350.095	1356.425	**120.9777**	**122.5229**	663.5429	668.5202
10t	1190.59	1196.753	593.9369	598.7746	**109.0402**	**110.5854**
11t	705.7238	710.8316	**76.2579**	**76.8632**	353.7721	357.4294
59t	1272.671	1278.977	648.9232	653.9449	**133.6733**	**135.5622**
61t	489.417	493.914	**98.1046**	**99.5207**	1052.214	1058.139
03t	1262.911	1268.985	615.0833	619.7448	**134.3924**	**135.9375**
50t	**111.6619**	**113.2071**	615.7123	620.55	1273.664	1279.888
52t	1417.828	1424.214	664.3024	669.2342	**142.8756**	**144.7644**
53t	1264.164	1270.376	**117.9222**	**119.3383**	641.491	646.3997
54t	1206.856	1212.955	**127.9285**	**129.595**	608.2039	612.9928

Table 4.19 Summary of the results of goodness-of-fit tests of GEV and GPD for $PM_{2.5}$ and PM_{10}.

	$PM_{2.5}$		PM_{10}	
Distribution	Non-rejected	Rejected	Non-rejected	Rejected
GEV	9	1	9	1
GPD (Quantile 90%)	4	6	10	0
GPD (Quantile 95%)	10	0	10	0
GPD (Quantile 99%)	10	0	9	1

Table 4.20 Summary of the results of the information criteria, AIC and BIC of GPD for $PM_{2.5}$ and PM_{10}.

	$PM_{2.5}$		PM_{10}	
Distribution	AIC	BIC	AIC	BIC
GPD (Quantile 90%)	0	0	1	1
GPD (Quantile 95%)	2	2	5	5
GPD (Quantile 99%)	8	8	4	4

Table 4.21 The results of return level of GPD Quantile 90% for $PM_{2.5}$.

$PM_{2.5}$	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	85.57301	93.16086	98.48333	101.4403
10t	73.40842	77.82278	80.64020	82.10757
11t	74.52877	78.90510	81.71650	83.18727
59t	78.04785	84.26333	88.44208	90.69727
61t	84.99790	90.85719	94.62904	96.60507
03t	182.9251	196.3067	205.5822	210.6935
50t	86.04794	92.37518	96.67471	99.01208
52t	98.6198	107.3891	113.5015	116.8827
53t	83.84076	87.04989	90.48671	91.76249
54t	104.3697	115.8876	124.6014	129.6986

Table 4.22 The results of return level of GPD Quantile 95% for $PM_{2.5}$.

$PM_{2.5}$	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	85.98631	92.80166	97.35523	99.80225
10t	73.47604	78.03232	80.96180	82.49524
11t	74.75324	78.18903	80.19605	81.17986
59t	75.73331	83.91666	90.10712	93.72831
61t	85.24463	91.26886	95.15576	97.19524
03t	183.5711	197.7225	207.6535	213.1724
50t	85.26032	89.51717	92.04213	93.29287
52t	97.03063	102.6483	106.0281	107.7187
53t	83.47497	86.62222	88.38330	89.22149
54t	101.2360	107.8173	111.9911	114.1555

Table 4.23 The results of return level of GPD Quantile 99% for $PM_{2.5}$.

$PM_{2.5}$	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	85.34499	95.16075	102.5167	106.7922
10t	72.72518	78.01690	81.73854	83.80984
11t	73.62482	79.49671	83.93863	86.53698
59t	77.98710	85.94687	91.49857	94.57089
61t	84.37368	92.00207	97.77272	101.1483
03t	186.7035	200.8517	209.7522	214.3421
50t	85.45971	89.35462	91.59935	92.68957
52t	98.47095	102.1640	103.8509	104.5494
53t	84.22529	86.74340	87.95850	88.48097
54t	96.56881	104.2202	110.8041	113.3941

Table 4.24 The results of return level of GPD Quantile 90% for PM_{10}.

PM_{10}	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	28.34050	137.5214	143.6973	147.0316
10t	115.9710	122.1572	126.1104	128.1710
11t	128.4566	141.8289	151.9447	157.8620
59t	108.3479	116.4416	121.9451	124.9383
61t	117.6074	122.0243	124.6525	125.9572
03t	182.9251	196.3067	205.5822	210.6935
50t	134.4538	140.8855	144.9085	146.9750
52t	142.3922	157.0017	168.0534	174.5182
53t	143.3459	153.6251	160.7482	164.6728
54t	154.5155	163.0256	168.6434	171.6375

Table 4.25 The results of return level of GPD Quantile 95% for PM_{10}.

PM_{10}	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	128.4303	139.0086	146.4811	150.6525
10t	115.9486	121.9944	125.8270	127.8138
11t	125.8589	133.8626	138.9681	141.6262
59t	108.6148	118.0207	124.7442	128.5279
61t	117.6838	122.1941	124.8830	126.2196
03t	183.5711	197.7225	207.6535	213.1724
50t	134.6471	141.0399	145.0064	147.0326
52t	142.8515	154.5373	162.8077	167.4305
53t	142.9403	154.1015	162.0944	166.5982
54t	155.1139	164.7393	171.3301	174.9318

Table 4.26 The results of return level of GPD Quantile 99% for PM_{10}.

PM_{10}	Return levels			
Station code	2-year	5-year	10-year	15-year
05t	128.9911	140.0214	147.5602	151.6748
10t	117.7721	121.4044	123.0199	123.6767
11t	126.8068	134.6925	139.8885	141.6068
59t	108.3799	118.8343	126.8370	131.5564
61t	118.3664	122.5514	124.7771	125.8002
03t	186.7035	200.8517	209.7522	214.3421
50t	135.2742	141.0923	144.3424	145.8875
52t	142.5156	155.3446	165.0494	170.7263
53t	130.9651	137.4471	139.7137	140.2745
54t	157.3435	164.2398	167.9470	169.6636

CHAPTER V
CONCLUSION

This research focuses on the daily average concentrations of $PM_{2.5}$ and PM_{10} categorized by years, specifically from 2018 to 2022, spanning the period from January 1, 2018, to December 31, 2022, for the ten previously mentioned stations in Bangkok The result of the non-mixture distribution analysis indicates that the concentrations of $PM_{2.5}$ and PM_{10} are most consistent with the lognormal distribution compared to other statistical distributions.

Furthermore, the study of mixture distributions are important because ozone, carbon monoxide, sulfur dioxide, nitrogen oxides, nitric oxide, and nitrogen dioxide are major components of $PM_{2.5}$ and PM_{10}. Taylor, Jakeman, and Simpson (1986) found that these air pollutants have log-normal, gamma, and Weibull distributions. These components come from various sources. Therefore, studying mixture distributions are crucial for understanding the complexity of the data, as it may comprise two or more components with different distribution characteristics. As a result, 2-mixture lognormal distribution is found to be the most consistent with the data for the 2-mixture distribution. For result of the 3-mixture distribution, 3-mixture lognormal distribution and the gamma-gamma-lognormal distribution are found to be the most consistent with the data for PM_{10} and $PM_{2.5}$, respectively.

Nevertheless, understanding the distribution of pollution alone does not fully address the issue for the public or the government. Being aware of trends in maximum pollution levels, characterized by return levels and return periods, can help raise public awareness of pollution issues. Therefore, we have chosen to investigate the extreme value analysis, namely the generalized extreme value distribution and the generalized Pareto distribution. Data for these two distributions varies slightly as mentioned earlier.

This research has found that the concentrations of $PM_{2.5}$ and PM_{10} are most consistent with the generalized extreme value distribution and the generalized Pareto distribution with a selected threshold from quantile 99%. However, caution is warranted for the generalized Pareto distribution with a selected threshold from quantile 99%, as it may lead to overfitting.

When examining the return levels of $PM_{2.5}$ and PM_{10}, it was observed that they consistently increased each year. This trend can be attributed to several factors, such as the growing usage of energy sources that are primary contributors to $PM_{2.5}$ and PM_{10} emissions, the increased production and utilization of automobiles, which emit pollutants contributing to $PM_{2.5}$ and PM_{10}, as well as the expansion of construction and industrial activities releasing fine particulate matter into the atmosphere. Additionally, fluctuations in weather patterns and changes in climatic conditions may also contribute to the annual escalation of $PM_{2.5}$ and PM_{10} levels. Effective control and management of these factors could potentially mitigate the rise in $PM_{2.5}$ and PM_{10} concentrations in the future.

REFERENCES

REFERENCES

Abild, J., Andersen, E. Y., and Rosbjerg, D. (1992). The climate of extreme winds at the great belt, Denmark. *Journal of Wind Engineering and Industrial Aerodynamics*, *41*(1-3), 521-532. https://doi.org/10.1016/0167-6105(92)90458-M.

Almeida, S. M., Pio, C. A., Freitas, M. C., Reis, M. A., and Trancoso, M. A. (2006). Approaching $PM_{2.5}$ and $PM_{2.5-10}$ source apportionment by mass balance analysis, principal component analysis and particle size distribution. *Science of The Total Environment*, *368*(2-3), 663-674. https://doi.org/10.1016/j.scitotenv.2006.03.031.

Chu, H. J., Yu, H. L., and Kuo, Y. M. (2012). Identifying spatial mixture distributions of $PM_{2.5}$ and PM_{10} in Taiwan during and after a dust storm. *Atmospheric Environment*, *54*, 728-737. https://doi.org/10.1016/j.atmosenv.2012.01.022.

Chuersuwan, N., Nimrat, S., Lekphet, S., and Kerdkumrai, T. (2008). Levels and major sources of $PM_{2.5}$ and PM_{10} in Bangkok Metropolitan Region. *Environment International*, *34*(5), 671-677. https://doi.org/10.1016/j.envint.2007.12.018.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Ding, J., Tarokh, V., and Yang, Y. (2018). Model Selection Techniques: An Overview, *IEEE Signal Processing Magazine*, *35*(6), 16-34. https://doi.org/10.1109/MSP.2018.2867638.

Erickson, L. E., Griswold, W., Maghirang, R. G., and Urbaszewski, B. P. (2016). The economic consequences of outdoor air pollution. *Journal of Environmental Protection*, *8*(10), OECD Publishing, Paris, France.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2019). *Introduction to Mathematical Statistics*. Pearson Education.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Taylor and Francis Group.

Kvam, P. H., and Vidakovic, B. (2007). *Nonparametric Statistics with Applications to Science and Engineering*. Wiley-Interscience.

Lu, H. C. (2004). Estimating the emission source reduction of PM_{10} in central Taiwan. *Chemosphere*, *54*(7), 805-814. https://doi.org/10.1016/j.chemosphere.2003.10.012.

Mage, D. M., and Ott, W. R. (1984). An evaluation of the methods of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. *Atmospheric Environment (1967)*, *18*(1), 163-171. https://doi.org/10.1016/0004-6981(84)90239-7.

Martins, L. D., Wikuats, C. F. H., Capucim, M. N., Almeida, D. S., Costa, S. C., Albuquerque, T., Carvalho, V. S. B., Freitas, E. D., Andrade, M. F., and Martins, J. A. (2017). Extreme value analysis of air pollution data and their comparison between two large urban regions of South America. *Weather and Climate Extremes*, *18*, 44-54. https://doi.org/10.1016/j.wace.2017.10.004.

Mijić, Z., Tasić, M., Rajšić, S., and Novaković, V. (2009). The statistical characters of PM_{10} in Belgrade area. *Atmospheric Research*, *92*(4), 420-426. https://doi.org/10.1016/j.atmosres.2009.01.002.

Mishara, G., Ghosh, K., Dwivedi, A. K., Kumar, M., Kumar, S. Chintalapati, S., and Tripathi, S, N. (2021). An application of probability density function for the analysis of $PM_{2.5}$ concentration during the COVID-19 lockdown period, Taiwan. *Science of The Total Environment*, *782*, 1-10. https://doi.org/10.1016/j.scitotenv.2021.146681.

Plocoste, T., Calif, R., Rajšić, S., Clotilde, L. E., and Brute, F. N. (2020). The statistical behavior of PM_{10} events over Guadeloupean archipelago: Stationarity, modelling and extreme events. *Atmospheric Research*, *241*, 1-21. https://doi.org/10.1016/j.atmosres.2020.104956.

Pollution Control Department. (2023, September 14). *Report on The Situation and Management of Air and Noise Pollution Problems in Thailand*. Pollution Control Department. Retrieved from https://www.pcd.go.th/publication/30447.

Rakhecha, P. R., and Singh, V. P. (2009). *Applied Hydrometeorology*. Springer.

Shi, Y., Matsunaga, T., Yamaguchi, Y., Zhao, A., Li, Z., and Gu, X. (2018). Long-term trends and spatial patterns of $PM_{2.5}$- induced premature mortality in South and Southeast Asia from 1999 to 2014. *Science of The Total Environment*, *631–632*, 1504-1514. https://doi.org/10.1016/j.scitotenv.2018.03.146.

Taylor, J. A., Jakeman, A. J., and Simpson, R. W. (1986). Modeling distributions of air pollutant concentrations—I. Identification of statistical models. *Atmospheric Environment (1967)*, *20*(9), 1781-1789. https://doi.org/10.1016/0004-6981(86)90127-7.

Walck, C. (2007). *Handbook on Statistical Distributions for Experimentalists*. Particle Physics Group, Fysikum University of Stockholm, Stockholm.

WHO. (2021). Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon dioxide. *WHO Global Air Quality Guideline*.

Xi, W., Ren J. C., Bing, H. C., and Hai, D. K. (2013). Taiwan PM_{10} concentration in air quality management in 5 representative cities of China. *Biomedical and Environmental Sciences*, *26*(8), 638-646. https://doi.org/10.3967/0895-3988.2013.08.002.

APPENDICES

APPENDIX A

RESULT FOR NON-MIXTURE DISTRIBUTION AND MIXTURE DISTRIBUTIONS

The statistics and p-values for goodness-of-fit, as well as the information criteria by non-mixture distribution and mixture distributions are presented in this chapter.

A.1 Non-Mixture Distributions for $PM_{2.5}$

Table A.1 Results of the goodness-of-fit tests (KS, CM, AD) and the iformation criteria (AIC and BIC) of gamma distribution for $PM_{2.5}$ in 2018. Non-rejections above the 5% level are marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11030 (0.00040)	0.90061 (0.00419)	5.02218 (0.00280)	2680.40	2688.09
10t	0.10970 **(0.3864)**	0.17643 **(0.3187)**	1.13130 **(0.29500)**	525.803	530.242
11t	0.09674 **(0.4754)**	4.1312 (2.20E-16)	0.9166 **(0.40380)**	567.449	572.111
59t	0.11357 (0.00087)	0.84749 (0.00559)	4.9499 (0.00304)	2218.12	2225.53
61t	0.16496 (0.5.77E-08)	1.50495 (0.000164)	8.24273 (8.62E-05)	2476.62	2484.15
03t	0.05737 **(0.9638)**	0.02432 **(0.9913)**	0.1623 **(0.9975)**	629.1192	633.7807
50t	0.11218 (0.00022)	1.02379 (0.002144)	5.93079 (0.00104)	2786.862	2794.645
52t	0.130101 (0.1.05E-05)	1.57713 (0.00011)	8.82979 (4.28E-05)	2933.018	2940.784
53t	0.11435 (0.0001431)	0.79482 (0.00749)	4.62587 (0.00434)	2830.767	2838.567
54t	0.04791 **(0.8639)**	0.04349 **(0.9149)**	0.38141 (0.8668)	1230.787	1236.9

Table A.2 Results of the goodness-of-fit tests (KS, CM, AD) and the iformation criteria (AIC and BIC) of lognormal distribution for $PM_{2.5}$ in 2018. Non-rejections above the 5% level are marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.07086 **(0.06192)**	0.31652 **(0.1215)**	1.88695 **(0.1061)**	2648.845	2656.538
10t	0.08879 **(0.6572)**	0.97871 **(0.5968)**	0.70691 **(0.5525)**	523.6511	528.0902
11t	0.08363 **(0.6625)**	0.07476 **(0.7242)**	0.53079 **(0.7151)**	563.2346	567.896
59t	0.07456 **(0.07121)**	0.28174 **(0.1523)**	1.7586 **(0.1252)**	2180.343	2187.751
61t	0.13116 (3.43E-05)	0.77411 (0.00839)	4.30014 (0.00624)	2434.407	2441.937
03t	0.06009 **(0.9466)**	0.02459 **(0.9908)**	0.16554 **(0.9971)**	629.1286	633.7901
50t	0.094734 (0.00301)	0.56153 (0.02783)	3.29069 (0.01955)	2752.017	2759.8
52t	0.10457 (0.00078)	0.81265 (0.00679)	4.70627 (0.00397)	2886.228	2893.995
53t	0.09533 (0.00263)	0.40207 **(0.07129)**	2.30354 **(0.0629)**	2801.831	2809.631
54t	0.05424 **(0.7448)**	0.05733 **(0.8315)**	0.41019 **(0.8384)**	1234.619	1240.732

Table A.3 Results of the goodness-of-fit tests (KS, CM, AD) and the iformation criteria (AIC and BIC) of Weibull distribution for $PM_{2.5}$ in 2018. Non-rejections above the 5% level are marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11713 (0.00015)	1.25535 (0.000618)	7.52763 (0.00019)	2718.035	2725.728
10t	0.13585 **(0.1625)**	0.32882 **(0.1122)**	2.02769 **(0.08887)**	535.0416	539.4806
11t	0.12644 **(0.176)**	0.30186 **(0.1335)**	2.0556 **(0.08579)**	581.4973	586.1588
59t	0.12749 (0.00012)	1.4217 (0.00026)	8.7534 (4.73E-05)	2266.865	2187.751
61t	0.16935 (2.26E-08)	2.08026 (8.02E-06)	11.97475 (1.90E-06)	2530.048	2537.578
03t	0.07329 **(0.8089)**	0.08272 **(0.6776)**	0.5957 **0.6515)**	635.8376	640.499
50t	0.12479 (2.54E-05)	1.61266 (9.33E-05)	10.11165 (7.15E-06)	2851.858	2859.642
52t	0.13763 (2.48E-06)	1.94926 (1.60E-05)	11.52743 (1.80E-06)	2980.939	2988.706
53t	0.10799 (0.00040)	1.18893 (0.00088)	7.56002 (0.00018)	2878.791	2886.59
54t	0.07512 **(0.3383)**	0.27122 **(0.1633)**	2.13939 **(0.07718)**	1259.824	1265.937

Table A.4 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for $PM_{2.5}$ in 2019. Non-rejections above the 5% level are marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.08766 (0.00744)	0.52116 (0.03515)	2.98081 (0.02801)	2767.88	2775.67
10t	0.07029 **(0.06108)**	0.27108 **(0.1634)**	1.88841 **(0.1059)**	2661.03	2668.76
11t	0.08091 (0.01796)	1.24800 (0.00064)	3.50095 (0.01537)	2707.95	2715.73
59t	0.08574 (0.00949)	0.56048 (0.02800)	3.45665 (0.01616)	2703.55	2711.34
61t	0.06476 **(0.09931)**	0.36031 **(0.09216)**	2.44867 **(0.05274)**	2770.38	2778.15
03t	0.09721 (0.00226)	0.57504 (0.02575)	3.88543 (0.00993)	2818.65	2826.41
50t	0.10291 (0.00092)	0.70328 (0.01246)	4.31546 (0.00613)	2768.75	2776.54
52t	0.08806 (0.00696)	0.67387 (0.01469)	4.06869 (0.00808)	2852.76	2860.56
53t	0.07631 (0.02851)	0.48147 (0.04435)	2.84758 (0.03275)	2823.96	2831.76
54t	0.08634 (0.00906)	0.49959 (0.03987)	3.01665 (0.02686)	2815.09	2822.87

Table A.5 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for $PM_{2.5}$ in 2019. Non-rejections above the 5% level are marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.05176 **(0.28370)**	0.12389 **(0.48010)**	0.77129 **(0.50210)**	2747.04	2754.84
10t	0.04358 **(0.51400)**	0.06958 **(0.75450)**	0.50728 **(0.73930)**	2640.07	2647.80
11t	0.06827 **(0.06975)**	117.940 (2.20E-16)	1.50859 **(0.17440)**	2681.82	2689.59
59t	0.06768 **(0.07124)**	0.21605 **(0.23850)**	1.39348 **(0.20410)**	2679.39	2687.19
61t	0.04542 **(0.45130)**	0.07557 **(0.71820)**	0.56390 **(0.68260)**	2744.39	2752.16
03t	0.07178 (0.04950)	0.25009 **(0.18830)**	1.75055 **(0.12660)**	2791.62	2799.38
50t	0.07406 (0.03731)	0.28306 **(0.15100)**	1.78856 **(0.12050)**	2735.85	2743.64
52t	0.05492 **(0.22090)**	0.17626 **(0.31870)**	1.17389 **(0.27770)**	2820.82	2828.62
53t	0.05237 **(0.26950)**	0.16310 **(0.35210)**	0.97291 **(0.37180)**	2805.84	2813.64
54t	0.06351 **(0.10790)**	0.20460 **(0.25880)**	1.27107 **(0.24200)**	2790.38	2798.16

Table A.6 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for $PM_{2.5}$ in 2019. Non-rejections above the 5% level are marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09877 (0.00165)	0.89096 (0.00441)	5.49251 (0.00167)	2801.31	2809.10
10t	0.08421 (0.01339)	0.62177 (0.01972)	4.54203 (0.00476)	2698.19	2705.92
11t	0.09926 (0.00166)	0.98121 (0.00270)	6.60439 (0.00051)	2755.04	2762.81
59t	0.09589 (0.00248)	0.92009 (0.00377)	6.04872 (0.00092)	2742.99	2750.79
61t	0.09241 (0.00442)	0.91331 (0.00391)	6.36867 (0.00065)	2820.66	2828.43
03t	0.13349 (5.55E-06)	1.80167 (3.46E-05)	11.57531 (1.78E-06)	2900.19	2907.96
50t	0.11437 (0.00015)	1.43842 (0.00023)	9.34687 (2.17E-05)	2835.63	2843.42
52t	0.10430 (0.00071)	1.18466 (0.00090)	7.57837 (0.00018)	2898.08	2905.88
53t	0.10993 (0.00029)	1.21857 (0.00075)	7.50632 (0.00019)	2872.99	2880.79
54t	0.10974 (0.00033)	1.23917 (0.00067)	8.20896 (8.94E-05)	2884.98	2892.76

Table A.7 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for $PM_{2.5}$ in 2020. Non-rejections above the 5% level are marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.13301 (4.75E-06)	1.42342 (0.00025)	7.80092 (0.00014)	2758.79	2766.59
10t	0.11123 (0.00029)	0.93483 (0.00347)	5.31373 (0.00204)	2608.71	2616.46
11t	0.12319 (3.18E-05)	1.23250 (0.00069)	7.18575 (0.00027)	2725.09	2732.88
59t	0.12114 (4.45E-05)	0.93520 (0.00347)	5.00735 (0.00285)	2719.77	2727.56
61t	0.11781 (0.00014)	1.46490 (0.00020)	8.57603 (5.83E-05)	2571.26	2578.95
03t	0.17011 (1.42E-09)	2.44624 (1.21E-06)	12.81644 (1.65E-06)	2849.47	2857.26
50t	0.13513 (3.76E-06)	1.37581 (0.00033)	7.78039 (0.00014)	2750.71	2758.48
52t	0.12843 (1.30E-05)	1.48502 (0.00018)	7.99427 (0.00011)	2768.28	2776.06
53t	0.10519 (0.00071)	0.73597 (0.01038)	4.09039 (0.00789)	2813.27	2821.04
54t	0.14194 (9.24E-07)	1.95732 (1.53E-05)	10.93092 (2.49E-06)	2864.02	2871.81

Table A.8 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for $PM_{2.5}$ in 2020. Non-rejections above the 5% level are marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09386 (0.00317)	0.63919 (0.01787)	3.54441 (0.01462)	2712.73	2720.54
10t	0.07955 (0.02182)	0.38938 **(0.07703)**	2.21956 **(0.06982)**	2573.12	2580.88
11t	0.09806 (0.00182)	0.62613 (0.00069)	3.81830 (0.01072)	2687.79	2695.58
59t	0.09277 (0.00374)	0.35144 **(0.09740)**	1.88903 **(0.10580)**	2691.77	2699.57
61t	0.10015 (0.00194)	0.75557 (0.00931)	4.61880 (0.00438)	2525.95	2533.64
03t	0.14967 (1.66E-07)	1.70198 (5.83E-05)	8.84012 (4.22E-05)	2805.67	2813.46
50t	0.10681 (0.00053)	0.73086 (0.01068)	4.27387 (0.00642)	2712.59	2720.37
52t	0.09358 (0.00353)	0.76122 (0.00902)	4.22995 (0.00675)	2730.17	2737.96
53t	0.07499 (0.03524)	0.30777 **(0.12850)**	1.83517 **(0.11340)**	2794.27	2802.04
54t	0.12804 (1.40E-05)	1.37822 (0.00032)	7.74311 (0.00015)	2819.82	2827.60

Table A.9 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for $PM_{2.5}$ in 2020. Non-rejections above the 5% level are marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.13131 (6.60E-06)	1.68359 (6.43E-05)	9.91819 (9.53E-06)	2797.68	2805.49
10t	0.12443 (3.16E-05)	1.41549 (0.00026)	8.71396 (4.93E-05)	2658.08	2665.84
11t	0.13045 (8.33E-06)	1.58359 (0.00011)	9.64453 (1.43E-05)	2767.95	2775.74
59t	0.11478 (0.00013)	1.26852 (0.00058)	7.33009 (0.00023)	2752.81	2760.61
61t	0.13878 (3.26E-06)	2.07348 (8.34E-06)	12.48419 (1.74E-06)	2629.33	2637.03
03t	0.16673 (3.25E-09)	3.12844 (3.53E-08)	17.20881 (1.65E-06)	2921.51	2929.31
50t	0.14417 (6.08E-07)	1.95961 (1.51E-05)	11.53026 (1.79E-06)	2807.71	2815.49
52t	0.13859 (1.82E-06)	1.89060 (2.17E-05)	10.76392 (2.95E-06)	2815.19	2822.98
53t	0.10891 (0.00040)	0.92121 (0.00374)	5.45153 (0.00175)	2836.25	2844.01
54t	0.15272 (9.28E-08)	2.56371 (6.55E-07)	15.1077 (1.66E-06)	2941.17	2948.97

Table A.10 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for $PM_{2.5}$ in 2021. Non-rejections above the 5% level are marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.10722 (0.00045)	0.76043 (0.00906)	4.47041 (0.00516)	2797.06	2804.86
10t	0.11714 (0.00012)	0.90903 (0.00399)	5.55597 (0.00156)	2659.58	2667.32
11t	0.13151 (6.57E-06)	1.08614 (0.00832)	6.42539 (0.00061)	2771.82	2779.62
59t	0.12129 (4.34E-05)	0.83996 (0.00584)	5.05048 (0.00272)	2722.69	2730.49
61t	0.14770 (3.15E-07)	2.37199 (1.77E-06)	14.1616 (1.67E-06)	2466.46	2474.22
03t	0.13957 (1.56E-06)	1.61203 (9.36E-05)	9.26565 (2.43E-05)	2878.32	2886.09
50t	0.12409 (2.97E-05)	1.04826 (0.00188)	6.43967 (0.00061)	2760.40	2768.18
52t	0.11708 (9.02E-05)	1.03682 (0.00199)	6.31084 (0.00069)	2813.85	2821.65
53t	0.09022 (0.00589)	0.44586 **(0.05476)**	2.77105 (0.03585)	2850.01	2857.77
54t	0.12180 (4.20E-05)	1.39075 (0.00030)	7.86487 (0.00013)	2887.57	2895.36

Table A.11 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for $PM_{2.5}$ in 2021. Non-rejections above the 5% level are marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.08496 (0.01030)	0.33791 **(0.10600)**	1.96160 **(0.09646)**	2763.93	2771.73
10t	0.09765 (0.00234)	0.50057 (0.03964)	3.12803 (0.02360)	2631.82	2639.56
11t	0.11468 (0.00014)	0.70105 (0.01262)	4.06017 (0.00816)	2737.43	2745.23
59t	0.10312 (0.00085)	0.45701 **(0.05125)**	2.55419 (0.04643)	2687.22	2695.02
61t	0.13762 (2.48E-06)	1.70031 (5.88E-05)	10.48176 (4.21E-06)	2416.36	2424.12
03t	0.12316 (3.51E-05)	1.03818 (0.00198)	5.84641 (0.00114)	2827.13	2834.91
50t	0.11298 (0.00019)	0.66923 (0.01508)	4.03153 (0.00843)	2723.85	2731.63
52t	0.10439 (0.00070)	0.59542 (0.02292)	3.54838 (0.01456)	2774.33	2782.13
53t	0.07203 (0.04872)	0.30698 **(0.12920)**	1.69882 **(0.13540)**	2832.40	2840.16
54t	0.10863 (0.00038)	0.85382 (0.00541)	4.89753 (0.00321)	2849.93	2857.72

Table A.12 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for $PM_{2.5}$ in 2021. Non-rejections above the 5% level are marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09593 (0.00242)	0.87118 (0.00492)	5.72009 (0.00131)	2824.49	2832.29
10t	0.11555 (0.00016)	1.10339 (0.00139)	7.02899 (0.00032)	2689.63	2697.36
11t	0.11681 (9.45E-05)	1.21459 (0.00077)	7.81029 (0.00014)	2807.68	2815.48
59t	0.10702 (0.00047)	0.97147 (0.00285)	6.51070 (0.00056)	2753.82	2761.62
61t	0.18939 (1.31E-11)	3.38177 (9.53E-09)	19.9739 (1.67E-06)	2559.89	2567.66
03t	0.14854 (2.42E-07)	2.26819 (3.04E-06)	14.07559 (1.66E-06)	2955.17	2962.95
50t	0.12342 (3.35E-05)	1.41688 (0.00026)	9.35153 (0.00002)	2814.82	2822.59
52t	0.11368 (0.00016)	1.45411 (0.00022)	9.50178 (1.75E-05)	2869.06	2876.86
53t	0.07899 (0.02297)	0.54534 (0.03055)	3.91641 (0.00959)	2874.73	2882.49
54t	0.13256 (5.76E-06)	2.01482 (1.14E-05)	12.07667 (1.67E-06)	2956.69	2964.48

Table A.13 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for $PM_{2.5}$ in 2022. Non-rejections above the 5% level are marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.12288 (3.47E-05)	0.91656 **(0.00384)**	5.27207 **(0.00213)**	2657.87	2665.66
10t	0.08517 (0.01340)	0.37831 (0.08245)	2.32999 (0.06095)	2428.10	2435.79
11t	0.09048 (0.00524)	0.54329 (2.20E-16)	3.27579 (0.01989)	2611.31	2619.09
59t	0.10837 (0.00039)	0.61961 (0.01997)	3.86828 (0.01013)	2528.69	2536.49
61t	0.14513 (4.57E-07)	1.76929 (4.10E-05)	10.60135 (3.58E-06)	2402.21	2409.99
03t	0.12225 (3.77E-05)	1.02486 (0.002132)	6.05756 (0.00091)	2677.29	2685.09
50t	0.13181 (8.49E-06)	1.23528 (0.00069)	7.10733 (0.00029)	2590.74	2598.49
52t	0.11975 (6.38E-05)	0.87119 (0.00492)	5.16909 (0.00239)	2650.09	2657.87
53t	0.10789 (0.00116)	0.77324 (0.00843)	4.44228 (0.00532)	2390.07	2397.60
54t	0.07342 (0.03993)	0.28848 **(0.14570)**	1.78119 **(0.12160)**	2678.10	2685.89

Table A.14 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for $PM_{2.5}$ in 2022. Non-rejections above the 5% level are marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.10385 (0.00079)	0.59433 (0.02306)	3.35858 (0.01809)	2632.13	2639.92
10t	0.06561 **(0.10260)**	0.20649 **(0.25530)**	1.29614 **(0.23360)**	2413.05	2420.74
11t	0.07382 (0.03825)	0.34720 **(0.10000)**	2.06561 **(0.08459)**	2591.06	2598.85
59t	0.09934 (0.00155)	0.45526 **(0.05179)**	2.70053 (0.03898)	2502.75	2510.54
61t	0.13688 (2.48E-06)	1.33488 (0.00041)	8.18893 (9.14E-05)	2363.09	2370.88
03t	0.11196 (0.00022)	0.68511 (0.01380)	4.00156 (0.00872)	2645.57	2653.36
50t	0.12097 (5.97E-05)	0.86067 (0.00521)	4.94252 (0.00306)	2562.94	2570.69
52t	0.10562 (0.000635)	0.59859 (0.02251)	3.41718 (0.01691)	2623.54	2631.32
53t	0.09078 (0.01024)	0.47689 (0.04555)	2.71219 (0.03845)	2367.71	2375.25
54t	0.05375 **(0.24520)**	0.16315 **(0.35200)**	0.97909 **(0.36840)**	2664.48	2672.27

Table A.15 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for $PM_{2.5}$ in 2022. Non-rejections above the 5% level are marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.12418 (2.75E-05)	1.26589 (0.00059)	7.99119 (0.00011)	2708.93	2716.72
10t	0.09319 (0.00499)	0.64340 (0.01745)	4.36727 (0.00579)	2464.08	2471.77
11t	0.09925 (0.00157)	0.77917 (0.00817)	5.33421 (0.00199)	2653.83	2661.62
59t	0.10359 (0.00083)	0.758029 (0.00918)	5.54166 (0.00159)	2570.18	2577.97
61t	0.17657 (2.96E-10)	2.55043 (7.02E-07)	15.65579 (1.65E-06)	2493.52	2501.31
03t	0.15045 (1.40E-07)	1.78378 (3.80E-05)	11.4927 (1.79E-06)	2764.08	2771.87
50t	0.13152 (8.97E-06)	1.75324 (4.46E-05)	10.6784 (3.28E-06)	2654.39	2662.15
52t	0.11551 (0.00013)	1.11044 (0.00134)	7.32023 (0.00024)	2696.74	2704.51
53t	0.11901 (0.00023)	0.96074 (0.00302)	6.04888 (0.00092)	2423.62	2431.16
54t	0.09186 (0.00437)	0.78317 (0.007998)	5.64251 (0.00142)	2738.47	2746.26

A.2 Non-Mixture Distributions for PM_{10}

Table A.16 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for PM_{10} in 2018. Non-rejections are the 5% level at above marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.08442 (0.01306)	0.76468 (0.00885)	4.77060 (0.00369)	3041.28	3049.01
10t	0.11498 (0.00091)	0.73696 (0.01031)	4.10720 (0.00774)	2490.75	2498.09
11t	0.13974 **(0.52020)**	0.30155 **(0.13360)**	0.45747 **(0.78930)**	274.012	277.065
59t	0.09228 (0.00434)	0.69654 (0.01294)	4.1503 (0.00738)	3007.29	3015.07
61t	0.11349 (0.00049)	1.31150 (0.00046)	7.78560 (0.00014)	2731.35	2738.89
03t	0.05189 **(0.28810)**	0.18224 **(0.30480)**	1.08300 **(0.31650)**	3221.59	3229.36
50t	0.08145 (0.01707)	0.59279 (0.02326)	3.62270 (0.01338)	3129.99	3137.76
52t	0.10343 (0.00092)	0.91158 (0.00394)	5.30910 (0.00205)	3249.00	3256.77
53t	.01006 (0.00124)	0.96714 (0.00291)	5.72730 (0.00129)	3284.76	3292.56
54t	0.06979 **(0.42490)**	0.08593 **(0.65900)**	0.48621 **(0.76070)**	1424.53	14300.66

Table A.17 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for PM_{10} in 2018. Non-rejections are the 5% level at above marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.06432 **(0.1078)**	0.30256 **(0.1329)**	1.9899 **(0.09306)**	3008.023	3015.756
10t	0.08840 (0.02117)	0.34378 **(0.10220)**	1.9198 **(0.10170)**	2469.81	2477.15
11t	0.13035 **(0.61030)**	0.07015 **(0.75390)**	0.40589 **(0.84190)**	273.422	276.475
59t	0.06828 **(0.06967)**	0.28618 **(0.14790)**	1.78580 **(0.12090)**	2976.37	2984.14
61t	0.09778 (0.00424)	0.73735 (0.01030)	4.55230 (0.00471)	2696.18	2703.73
03t	0.03651 **(0.72330)**	0.05449 **(0.84890)**	0.33531 **(0.90940)**	3214.09	3221.87
50t	0.06881 **(0.13970)**	0.28836 **(0.14580)**	1.83020 **(0.11410)**	3108.29	3116.05
52t	0.07735 (0.02724)	0.54219 (0.03112)	3.26860 (0.02006)	3222.99	32300.76
53t	0.08732 (0.00764)	0.57431 (0.02586)	3.15349 (0.01515)	3255.99	3263.79
54t	0.09343 **(0.12670)**	0.18623 **(0.29600)**	0.99082 **(0.36200)**	1427.90	1434.03

Table A.18 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for PM_{10} in 2018. Non-rejections are the 5% level at above marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.10799 (0.00053)	1.37213 (0.00033)	8.86498 (4.09E-05)	3097.72	3105.45
10t	0.12023 (0.00044)	1.18950 (0.00088)	7.08050 (0.00031)	2530.84	2538.18
11t	0.14732 **(0.45170)**	0.10997 **(0.54060)**	0.68040 **(0.57420)**	277.33	280.38
59t	0.10344 (0.00090)	1.20550 (0.00081)	7.72250 (0.00015)	3061.63	3069.39
61t	0.13523 (1.54E-05)	2.03210 (1.03E-05)	12.1490 (1.87E-06)	2792.18	2799.73
03t	0.09571 (0.00273)	0.86699 (0.00503)	5.3929 (0.00187)	3272.69	3280.47
50t	0.11062 (0.00031)	1.25000 (0.00064)	7.95290 (0.00012)	3191.21	3198.98
52t	0.11893 (7.77E-05)	1.25730 (0.00061)	7.79580 (0.00014)	3296.15	3303.92
53t	0.11383 (0.00016)	1.35875 (0.00036)	8.50392 (6.35E-05)	3336.85	3344.65
54t	0.04893 **(0.84370)**	0.07739 **(0.70780)**	0.69356 **(0.56400)**	1433.36	1439.48

Table A.19 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for PM_{10} in 2019. Non-rejections are the 5% level at above marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.07932 (0.02024)	0.45259 **(0.05261)**	2.57709 (0.04517)	3122.13	3129.93
10t	0.08134 (0.01754)	0.43456 **(0.05860)**	2.59910 (0.04400)	3000.87	3008.63
11t	0.08916 **(0.13400)**	0.96974 (0.00285)	1.51020 **(0.17410)**	1455.86	1462.13
59t	0.09373 (0.003281)	0.81831 (0.00658)	4.90500 (0.00319)	2964.97	2972.77
61t	0.075162 (0.03348)	0.50374 (0.03891)	3.3216 (0.01887)	3067.695	3075.478
03t	0.07529 (0.03340)	0.42836 **(0.06082)**	2.64970 (0.04142)	3227.79	3235.57
50t	0.07549 (0.03119)	0.50775 (0.03801)	3.10760 (0.02416)	3120.32	3128.12
52t	0.10628 (0.00056)	0.94661 (0.00326)	5.24590 (0.00219)	3135.31	3143.09
53t	0.04904 **(0.34540)**	0.21787 **(0.23540)**	1.64770 **(0.14480)**	3193.36	3201.16
54t	0.04222 **(0.54600)**	0.10691 **(0.55230)**	0.75842 **(0.51190)**	3115.41	3123.18

Table A.20 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for PM_{10} in 2019. Non-rejections are the 5% level at above marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.04825 **(0.36320)**	0.13225 **(0.44870)**	0.83687 **(0.45510)**	3103.73	3111.53
10t	0.04901 **(0.35610)**	0.14315 **(0.41130)**	0.97795 **(0.36900)**	2981.94	2989.69
11t	0.09322 **(0.10420)**	0.19454 **(0.27840)**	1.38020 **(0.20780)**	1455.67	1461.95
59t	0.08177 (0.01518)	0.44246 **(0.05589)**	2.69960 (0.03902)	2937.03	2944.83
61t	0.05793 **(0.17610)**	0.24463 **(0.19540)**	1.79380 **(0.11960)**	3046.76	3054.54
03t	0.05993 **(0.14950)**	0.17076 **(0.33220)**	1.0518 **(0.33120)**	3207.75	3215.53
50t	0.05168 **(0.28380)**	0.19983 **(0.26790)**	1.26110 **(0.24540)**	3098.51	3106.31
52t	0.07628 **(0.02962)**	0.45465 **(0.05197)**	2.54910 (0.04672)	3107.20	3114.99
53t	0.04359 **(0.49360)**	0.07923 **(0.69660)**	0.68283 **(0.57320)**	3178.60	3186.39
54t	0.03579 **(0.74860)**	0.06274 **(0.79710)**	0.37736 **(0.87080)**	3108.48	3116.24

Table A.21 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for PM_{10} in 2019. Non-rejections are the 5% level at above marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11064 (0.00026)	0.96421 (0.00296)	5.80404 (0.00119)	3164.79	3172.59
10t	0.09602 (0.00272)	0.82247 (0.00643)	5.17680 (0.00237)	3039.84	3047.59
11t	0.07909 **(0.23800)**	0.29704 **(0.13780)**	1.92100 **(0.10160)**	1465.28	1471.55
59t	0.11007 (0.00029)	1.29200 (0.00051)	8.22980 (8.72E-05)	3018.90	3026.70
61t	0.10015 (0.00140)	1.00600 (0.00236)	6.65800 (0.00048)	3120.61	3128.39
03t	0.11373 (0.00018)	1.35540 (0.00036)	8.76420 (4.64E-05)	3301.90	3309.68
50t	0.11555 (0.00012)	1.40910 (0.00027)	9.01070 (3.40E-05)	3192.17	3199.97
52t	0.12486 (2.51E-05)	1.55270 (0.00013)	9.16420 (2.78E-05)	3189.63	3197.42
53t	0.07994 (0.01908)	0.62494 (0.01937)	4.47390 (0.005139)	3234.81	3242.60
54t	0.07756 (0.02694)	0.64384 (0.01741)	4.72210 (0.00390)	3171.14	3178.89

Table A.22 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for PM_{10} in 2020. Non-rejections are the 5% level at above marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.10442 (0.000698)	1.05276 (0.00183)	6.07409 (0.00089)	3080.89	3088.69
10t	0.11708 (0.00013)	1.28345 (0.00053)	7.66362 (0.00016)	2962.48	2970.21
11t	0.14471 (4.41E-07)	1.48808 (0.00013)	8.81102 (4.37E-05)	3078.94	3086.75
59t	0.09478 (0.00279)	0.67989 (0.01421)	3.85990 (0.01022)	2965.49	2973.29
61t	0.09931 (0.00149)	1.01566 (0.00224)	6.37521 (0.00065)	3094.06	3101.86
03t	0.15955 (2.09E-08)	2.50988 (8.66E-07)	13.65246 (1.66E-06)	3225.21	3232.99
50t	0.10952 (0.00032)	1.14579 (0.00111)	6.78621 (0.00041)	3112.58	3120.37
52t	0.10697 (0.00046)	1.17244 (0.00096)	6.76034 (0.00043)	3125.04	3132.85
53t	0.12563 (2.25E-05)	1.07773 (0.00160)	5.86143 (0.00112)	3202.45	3210.23
54t	0.11597 (0.00011)	1.06233 (0.00174)	6.04774 (0.00092)	3290.58	3298.38

Table A.23 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for PM_{10} in 2020. Non-rejections are the 5% level at above marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.07969 (0.01940)	0.46895 (0.04775)	2.76921 (0.03593)	3046.93	3054.73
10t	0.09868 (0.00211)	0.70923 (0.01205)	4.40784 (0.00553)	2927.35	2935.08
11t	0.11315 (0.00017)	0.74405 (0.00992)	4.63585 (0.00429)	3028.43	3036.23
59t	0.06820 **(0.06641)**	0.25649 **(0.18030)**	1.52024 **(0.17170)**	2942.78	2950.58
61t	0.07903 (0.02095)	0.51053 (0.03740)	3.45205 (0.01625)	3062.32	3070.12
03t	0.13575 (3.33E-06)	1.86332 (2.51E-05)	10.2132 (6.16E-06)	3181.42	3189.19
50t	0.09047 (0.00517)	0.62842 (0.01899)	3.83595 (0.01050)	3081.06	3088.86
52t	0.09186 (0.00415)	0.63769 (0.01802)	3.80562 (0.01087)	3093.82	3101.63
53t	0.09635 (0.00246)	0.66686 (0.01529)	3.69164 (0.01237)	3182.57	3190.34
54t	0.09292 (0.00373)	0.70558 (0.01230)	3.99117 (0.00882)	3264.89	3272.69

Table A.24 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for PM_{10} in 2020. Non-rejections are the 5% level at above marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11794 (7.79E-05)	1.48379 (0.00018)	8.99932 (3.45E-05)	3125.13	3132.93
10t	0.12133 (6.31E-05)	1.68139 (6.49E-05)	10.26404 (5.76E-06)	3007.09	3014.82
11t	0.14272 (6.70E-07)	2.02903 (1.06E-05)	12.46595 (1.64E-06)	3137.13	3144.94
59t	0.10873 (0.00035)	1.18991 (0.00088)	7.12727 (0.00029)	3008.86	3016.67
61t	0.11344 (0.00017)	1.56357 (0.00012)	9.80189 (1.13E-05)	3145.30	3153.10
03t	0.17109 (1.33E-09)	3.34397 (1.16E-08)	18.7973 (1.66E-06)	3314.38	3322.16
50t	0.12826 (1.26E-05)	1.98844 (1.30E-05)	11.93552 (1.67E-06)	3181.36	3189.15
52t	0.12118 (4.30E-05)	1.66527 (7.08E-05)	9.94321 (9.18E-06)	3176.44	3184.25
53t	0.13634 (2.97E-06)	1.35547 (0.00036)	7.80231 (0.00014)	3235.54	3243.32
54t	0.12747 (1.46E-05)	1.57095 (0.00012)	9.61369 (1.49E-05)	3352.25	3360.05

Table A.25 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for PM_{10} in 2021. Non-rejections are the 5% level at above marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09089 (0.00481)	0.67382 (0.01470)	4.09133 (0.00788)	3090.14	3097.94
10t	0.11098 (0.00034)	0.82449 (0.00636)	4.67960 (0.00409)	3030.98	3038.71
11t	0.10656 (0.00051)	0.75171 (0.04562)	4.52319 (0.00486)	3089.99	3097.78
59t	0.12293 (3.24E-05)	1.24180 (0.00067)	7.55030 (0.00019)	2906.35	2914.15
61t	0.12871 (1.37E-05)	0.98548 (0.00264)	6.1166 (0.00085)	3037.17	3044.94
03t	0.12833 (1.73E-05)	1.0809 (0.001575)	7.0915 (0.00030)	3282.69	3290.43
50t	0.10977 (0.00032)	0.83427 (0.00603)	4.82890 (0.00347)	3126.55	3134.34
52t	0.11604 (0.00011)	0.98651 (0.00262)	5.81780 (0.00118)	3143.38	3151.18
53t	0.07517 (0.03498)	0.41098 **(0.06754)**	2.40320 **(0.05574)**	3204.96	3212.73
54t	0.09989 (0.00143)	0.80244 (0.00718)	4.71530 (0.00393)	3196.71	3204.49

Table A.26 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for PM_{10} in 2021. Non-rejections are the 5% level at above marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.06194 **(0.12150)**	0.23964 **(0.20220)**	1.48099 **(0.18110)**	3056.95	3064.75
10t	0.09997 (0.00172)	0.57578 (0.02564)	3.11550 (0.02394)	3008.067	3015.8
11t	0.08859 (0.00660)	0.51499 (0.03644)	2.95893 (0.02874)	3064.16	3071.95
59t	0.10775 (0.00041)	0.79982 (0.00729)	4.90230 (0.00319)	2868.79	2876.59
61t	0.11388 (0.00018)	0.68608 (0.01372)	4.19790 (0.00699)	3006.37	3014.14
03t	0.12389 (3.82E-05)	0.81564 (0.006675)	5.26890 (0.002138)	3248.82	3256.56
50t	0.09924 (0.00157)	0.53730 (0.03201)	2.97520 (0.0282)	3100.16	3107.94
52t	0.10281 (0.00089)	0.56822 (0.02678)	3.30590 (0.01922)	3109.55	3117.35
53t	0.05907 **(0.16440)**	0.25779 **(0.17870)**	1.49150 **(0.17850)**	3193.21	3200.97
54t	0.08516 (0.01033)	0.51579 (0.0363)	2.98770 (0.02779)	3173.06	3180.85

Table A.27 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for PM_{10} in 2021. Non-rejections are the 5% level at above marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09302 (0.00361)	1.10991 (0.00135)	7.28745 (0.00025)	3137.84	3145.64
10t	0.10284 (0.00114)	1.00181 (0.00241)	6.47948 (0.000579)	3070.82	3078.55
11t	0.10204 (0.00102)	0.94182 (0.00335)	6.42745 (0.00061)	3132.08	3139.88
59t	0.13385 (4.18E-06)	1.77730 (3.93E-05)	11.32200 (1.90E-06)	2972.06	2979.86
61t	0.12285 (3.93E-05)	1.34880 (0.0003761)	9.01100 (3.40E-05)	3093.96	3101.73
03t	0.13948 (2.09E-06)	1.55390 (0.00013)	10.59100 (3.67E-06)	3350.77	3358.51
50t	0.10409 (0.00077)	1.25740 (0.00061)	8.16670 (9.37E-05)	3184.59	3192.38
52t	0.11151 (0.000229)	1.49410 (0.00017)	9.49050 (1.78E-05)	3204.17	3211.96
53t	0.080243 (0.01990)	0.58036 (0.02498)	3.86120 (0.01021)	3233.47	3241.25
54t	0.11852 (7.45E-05)	1.47620 (0.00019)	9.58110 (1.56E-05)	3270.92	3278.71

Table A.28 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of gamma distribution for PM_{10} in 2022. Non-rejections are the 5% level at above marked in bold.

Gamma distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11229 (0.00038)	0.96505 (0.00295)	5.58380 (0.00152)	2691.33	2698.98
10t	0.08859 (0.01666)	0.41688 **(0.06516)**	2.63830 (0.04199)	2498.67	2506.11
11t	0.08825 **(0.98450)**	0.03183 **(0.13100)**	0.21607 **(0.98540)**	191.395	193.987
59t	0.10116 (0.00124)	0.67487 (0.01461)	4.45763 (0.00523)	2741.20	2748.98
61t	0.06958 **(0.68390)**	0.04399 **(0.91250)**	0.29056 **(0.94500)**	872.621	877.948
03t	0.07559 **(0.59220)**	0.09732 **(0.59910)**	0.58086 **(0.66590)**	908.973	914.262
50t	0.10068 (0.01542)	0.53391 (0.03262)	3.21448 (0.02136)	1934.11	1941.07
52t	0.12172 (6.08E-05)	0.92641 (0.00364)	5.60275 (0.00149)	2802.58	2810.30
53t	0.08433 (0.02726)	0.40609 **(0.06956)**	2.41474 **(0.05497)**	2469.19	2476.62
54t	0.06384 **(0.10210)**	0.30616 **(0.12990)**	1.95495 **(0.09728)**	3079.60	3087.40

Table A.29 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of lognormal distribution for PM_{10} in 2022. Non-rejections are the 5% level at above marked in bold.

Lognormal distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.09389 (0.00498)	0.59957 (0.02238)	3.45940 (0.01612)	2664.66	2672.32
10t	0.08368 (0.02792)	0.31255 **(0.12460)**	1.92777 **(0.10070)**	2486.27	2493.71
11t	0.09941 **(0.95230)**	0.04338 **(0.91850)**	0.28438 **(0.94900)**	192.141	194.733
59t	0.09244 (0.00418)	0.49342 (0.04134)	3.32959 (0.01870)	2719.23	2727.00
61t	0.07942 **(0.51570)**	0.06745 **(0.76840)**	0.43836 **(0.80970)**	873.984	879.311
03t	0.05989 **(0.84970)**	0.05529 **(0.84470)**	0.36692 **(0.88050)**	908.276	913.565
50t	0.08205 **(0.07901)**	0.31475 **(0.12290)**	1.93287 **(0.10010)**	1917.72	1924.68
52t	0.11297 (0.00026)	0.65347 (0.01648)	3.85489 (0.01028)	2777.14	2784.87
53t	0.06308 **(0.18060)**	0.18526 **(0.29810)**	1.12037 **(0.29980)**	2455.81	2463.23
54t	0.04930 **(0.33750)**	0.16415 **(0.34930)**	1.09299 **(0.31190)**	3069.05	3076.85

Table A.30 Results of the goodness-of-fit tests (KS, CM, AD) and the information criteria (AIC and BIC) of Weibull distribution for PM_{10} in 2022. Non-rejections are the 5% level at above marked in bold.

Weibull distribution					
Station code	KS (p-value)	CM (p-value)	AD (p-value)	AIC	BIC
05t	0.11721 (0.00018)	1.5575 (0.00013)	9.6115 (1.51E-05)	2754.12	2761.78
10t	0.08034 (0.03901)	0.55847 (0.02832)	3.89498 (0.00983)	2527.34	2534.78
11t	0.06195 **(0.99990)**	0.016184 **(0.99950)**	0.13703 **(0.99940)**	190.943	193.535
59t	0.12142 (4.77E-05)	1.04629 (0.00189)	7.32896 (0.00024)	2804.03	2811.81
61t	0.06922 **(0.69010)**	0.07279 **(0.73560)**	0.57437 **(0.67220)**	878.379	883.706
03t	0.12053 **(0.09740)**	0.35590 **(0.09463)**	2.06818 **(0.08440)**	923.627	928.916
50t	0.12253 (0.00148)	1.07715 (0.00159)	6.79569 (0.00042)	1984.94	1991.89
52t	0.11613 (0.00016)	1.33290 (0.00041)	8.68122 (5.14E-05)	2859.35	2867.08
53t	0.09691 (0.00688)	0.83289 (0.00606)	5.25899 (0.00216)	2507.92	2515.34
54t	0.08858 (0.00651)	0.91626 (0.00385)	6.01459 (0.00095)	3138.59	3146.39

A.3 2-Mixture Distributions for $PM_{2.5}$

Table A.31 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of $PM_{2.5}$ in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of $PM_{2.5}$ in 2018						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2636.64	2637.95	2646.82	**2636.60**	NA	NA
10t	521.910	520.560	523.820	**519.908**	NA	NA
11t	566.380	566.680	571.380	**564.119**	NA	NA
59t	2170.89	**2168.53**	2187.69	2168.69	NA	NA
61t	**2398.47**	2398.72	2412.89	2398.77	NA	NA
03t	634.540	**634.430**	637.010	634.476	NA	NA
50t	2725.44	**2720.80**	2758.92	2723.66	NA	NA
52t	2840.67	2838.43	2863.24	**2837.97**	NA	NA
53t	2784.17	2783.11	2805.43	**2782.64**	NA	NA
54t	**1195.03**	1235.09	1238.64	1195.38	NA	NA
BIC of $PM_{2.5}$ in 2018						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2655.87	2657.18	2666.05	**2655.84**	NA	NA
10t	533.010	531.660	534.920	**531.006**	NA	NA
11t	578.030	578.330	583.030	**575.774**	NA	NA
59t	2189.41	**2187.05**	2206.21	2187.21	NA	NA
61t	2417.54	**2417.29**	2431.72	2417.59	NA	NA
03t	**646.120**	646.800	648.660	646.290	NA	NA
50t	2744.90	**2740.26**	2778.38	2743.12	NA	NA
52t	2860.09	2857.84	2882.66	**2857.39**	NA	NA
53t	2803.67	2802.61	2824.93	**2802.14**	NA	NA
54t	**1210.31**	1250.37	1253.93	1210.66	NA	NA

Table A.32 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of $PM_{2.5}$ in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of $PM_{2.5}$ in 2019						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2744.59	2744.40	2751.95	**2744.24**	NA	NA
10t	2647.88	2643.38	2661.66	**2641.542**	NA	NA
11t	2676.53	**2671.27**	2701.86	2673.55	NA	NA
59t	2674.30	2673.11	2693.11	**2672.84**	NA	NA
61t	2749.09	2746.26	2765.44	**2745.05**	NA	NA
03t	2774.76	**2774.21**	2787.00	2774.85	NA	NA
50t	2728.23	**2722.98**	2758.65	2724.51	NA	NA
52t	2818.09	2817.18	2832.37	**2816.68**	NA	NA
53t	**2800.31**	2802.77	2803.80	2801.38	NA	NA
54t	2787.96	**2783.91**	2814.07	2786.07	NA	NA
BIC of $PM_{2.5}$ in 2019						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2764.08	2763.89	2771.43	**2763.57**	NA	NA
10t	2667.21	2662.72	2681.00	**2660.87**	NA	NA
11t	2696.02	**2690.92**	2721.32	2692.97	NA	NA
59t	2693.86	2692.68	2712.62	**2692.32**	NA	NA
61t	2768.78	2765.60	2784.87	**2764.44**	NA	NA
03t	2794.06	**2793.60**	2806.43	2793.62	NA	NA
50t	2747.74	**2742.66**	2778.16	2744.16	NA	NA
52t	2837.71	2838.19	2851.88	**2836.25**	NA	NA
53t	**2819.81**	2820.49	2823.33	2822.35	NA	NA
54t	2807.44	**2803.58**	2833.52	2805.53	NA	NA

Table A.33 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of $PM_{2.5}$ in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of $PM_{2.5}$ in 2020						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2682.51	2682.51	2696.98	**2682.45**	NA	NA
10t	2558.34	**2556.89**	2576.81	2556.92	NA	NA
11t	2653.17	2653.35	2674.11	**2650.43**	NA	NA
59t	**2678.12**	2679.66	2681.69	2679.14	NA	NA
61t	2487.96	2484.29	2514.88	**2483.75**	NA	NA
03t	**2718.12**	2735.26	2732.02	2718.28	NA	NA
50t	2673.44	2672.56	2692.85	**2672.04**	NA	NA
52t	**2692.65**	2693.32	2702.43	2694.55	NA	NA
53t	2774.93	2774.96	2780.65	**2774.23**	NA	NA
54t	2741.76	**2735.69**	2785.72	2740.40	NA	NA
BIC of $PM_{2.5}$ in 2020						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2702.02	2702.02	2716.49	**2701.96**	NA	NA
10t	2577.73	**2576.28**	2596.20	2576.31	NA	NA
11t	2672.66	2672.83	2693.59	**2669.92**	NA	NA
59t	**2697.62**	2699.16	2701.19	2698.64	NA	NA
61t	2507.20	2503.53	2534.11	**2502.98**	NA	NA
03t	**2737.61**	2754.74	2751.51	2737.76	NA	NA
50t	2692.88	2692.01	2712.30	**2691.49**	NA	NA
52t	**2712.10**	2712.78	2721.89	2714.01	NA	NA
53t	2794.35	2794.38	2800.06	**2794.38**	NA	NA
54t	2761.22	**2755.15**	2805.18	2759.85	NA	NA

Table A.34 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of $PM_{2.5}$ in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of $PM_{2.5}$ in 2021						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2749.01	**2745.77**	2768.08	2747.25	NA	NA
10t	2600.82	2600.91	2614.96	**2599.53**	NA	NA
11t	2690.96	**2688.18**	2713.98	2690.24	NA	NA
59t	2667.01	**2663.48**	2683.16	2667.39	NA	NA
61t	2323.98	**2307.30**	2390.37	2321.26	NA	NA
03t	2773.66	**2760.23**	2814.88	2773.77	NA	NA
50t	2683.25	**2676.03**	2717.53	2683.27	NA	NA
52t	2744.34	**2735.94**	2779.51	2744.09	NA	NA
53t	2818.73	**2817.97**	2832.40	2819.36	NA	NA
54t	2806.84	**2803.10**	2843.41	2804.78	NA	NA
BIC of $PM_{2.5}$ in 2021						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2768.51	**2765.27**	2787.58	2766.75	NA	NA
10t	2620.17	2620.25	2634.31	**2618.87**	NA	NA
11t	2710.46	**2707.68**	2733.48	2709.74	NA	NA
59t	2686.51	**2682.98**	2702.66	2686.89	NA	NA
61t	2343.40	**2326.72**	2409.79	2340.67	NA	NA
03t	2793.10	**2779.68**	2834.33	2793.22	NA	NA
50t	2702.69	**2695.48**	2736.97	2702.71	NA	NA
52t	2763.84	**2755.44**	2799.01	2763.58	NA	NA
53t	2838.13	**2837.37**	2851.80	2838.76	NA	NA
54t	2826.31	**2822.57**	2862.88	2824.25	NA	NA

Table A.35 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of $PM_{2.5}$ in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of $PM_{2.5}$ in 2022						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2598.14	**2597.02**	2621.16	2597.98	NA	NA
10t	2405.37	**2403.53**	2425.41	2404.71	NA	NA
11t	2573.59	**2570.39**	2601.21	2571.45	NA	NA
59t	2479.33	**2474.03**	2511.82	2474.67	NA	NA
61t	2282.65	**2269.44**	2350.09	2271.41	NA	NA
03t	2611.63	**2603.05**	2663.81	2603.27	NA	NA
50t	2513.66	**2512.65**	2544.04	2512.67	NA	NA
52t	2588.14	2588.12	2612.59	**2586.62**	NA	NA
53t	2340.48	**2340.23**	2357.16	2340.43	NA	NA
54t	2664.80	2662.28	2695.25	**2662.26**	NA	NA
BIC of $PM_{2.5}$ in 2022						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2617.61	**2616.49**	2640.63	2617.45	NA	NA
10t	2424.59	**2422.75**	2444.63	2423.93	NA	NA
11t	2593.06	**2589.86**	2620.68	2590.93	NA	NA
59t	2498.80	**2493.51**	2531.30	2494.15	NA	NA
61t	2302.13	**2288.91**	2369.57	2290.88	NA	NA
03t	2631.11	**2622.54**	2683.29	2622.76	NA	NA
50t	2533.04	**2532.02**	2563.41	2532.04	NA	NA
52t	2607.59	2607.57	2632.04	**2606.06**	NA	NA
53t	2359.32	**2359.07**	2376.00	2359.27	NA	NA
54t	2684.27	2681.75	2714.73	**2681.73**	NA	NA

A.4 2-Mixture Distributions for PM_{10}

Table A.36 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of PM_{10} in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of PM_{10} in 2018						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2996.36	**2992.19**	3022.74	2992.82	NA	NA
10t	**2456.15**	2457.45	2464.27	2456.61	NA	NA
11t	276.664	**276.496**	277.672	276.696	NA	NA
59t	2965.59	**2960.98**	2992.53	2962.50	NA	NA
61t	2656.19	2653.79	2679.62	**2652.79**	NA	NA
03t	3216.93	3219.75	3221.52	**3216.76**	NA	NA
50t	3095.17	3093.51	3119.21	**3093.46**	NA	NA
52t	3188.36	**3186.22**	3211.76	3186.29	NA	NA
53t	3220.57	**3217.51**	3251.14	3217.91	NA	NA
54t	1426.08	**1425.12**	1437.35	1425.22	NA	NA
BIC of PM_{10} in 2018						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3015.69	**3011.53**	3042.07	3012.15	NA	NA
10t	**2474.51**	2475.82	2482.64	2474.97	NA	NA
11t	284.296	**284.128**	285.304	284.327	NA	NA
59t	2985.02	**2980.41**	3011.96	2981.94	NA	NA
61t	2675.07	2672.67	2698.50	**2671.67**	NA	NA
03t	3236.36	3239.18	3240.95	**3236.19**	NA	NA
50t	3114.59	3112.92	3138.62	**3112.87**	NA	NA
52t	3207.78	**3205.64**	3231.17	3205.71	NA	NA
53t	3240.07	**3237.01**	3270.64	3237.41	NA	NA
54t	1441.39	**1440.43**	1449.60	1440.53	NA	NA

Table A.37 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of PM_{10} in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of PM_{10} in 2019						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3100.70	**3099.56**	3107.01	3100.09	NA	NA
10t	2978.20	**2976.05**	2991.73	2976.30	NA	NA
11t	1446.18	**1434.72**	1453.12	1444.92	NA	NA
59t	2913.37	**2910.21**	2943.75	2911.28	NA	NA
61t	3035.60	**3028.43**	3062.12	3032.32	NA	NA
03t	3205.19	3204.16	3226.49	**3203.95**	NA	NA
50t	3093.71	3102.51	3113.22	**3092.95**	NA	NA
52t	3084.50	**3084.38**	3096.69	3084.77	NA	NA
53t	3181.99	**3172.27**	3197.77	3179.71	NA	NA
54t	3115.62	**3113.64**	3130.42	3114.12	NA	NA
BIC of PM_{10} in 2019						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3120.19	**3119.06**	3126.51	3119.59	NA	NA
10t	2997.60	**2995.46**	3011.13	2995.71	NA	NA
11t	1461.86	**1450.40**	1468.80	1460.60	NA	NA
59t	2932.87	**2929.71**	2963.25	2930.78	NA	NA
61t	3055.06	**3047.89**	3081.58	3051.78	NA	NA
03t	3224.63	3223.61	3245.94	**3223.40**	NA	NA
50t	3113.20	3118.11	3132.72	**3112.45**	NA	NA
52t	3103.96	**3103.84**	3116.15	3104.23	NA	NA
53t	3201.48	**3191.75**	3217.26	3199.19	NA	NA
54t	3135.02	**3133.04**	3149.82	3133.52	NA	NA

Table A.38 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of PM_{10} in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of PM_{10} in 2020						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3024.09	3022.71	3041.92	**3021.63**	NA	NA
10t	2884.41	2882.88	2907.44	**2880.97**	NA	NA
11t	2990.14	**2981.82**	3024.45	2984.91	NA	NA
59t	2930.72	2929.51	2940.37	**2929.20**	NA	NA
61t	3032.52	3029.40	3056.39	**3027.82**	NA	NA
03t	3078.10	**3074.71**	3114.41	3077.07	NA	NA
50t	3048.02	3046.81	3070.53	**3046.01**	NA	NA
52t	3059.45	3058.14	3079.98	**3057.70**	NA	NA
53t	3137.16	3137.99	3138.51	**3136.47**	NA	NA
54t	3224.02	**3224.00**	3249.40	3224.13	NA	NA
BIC of PM_{10} in 2020						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3043.59	3042.21	3061.42	**3041.12**	NA	NA
10t	2903.73	2902.20	2926.76	**2900.28**	NA	NA
11t	3009.66	**3001.34**	3043.96	3004.42	NA	NA
59t	2950.23	2949.02	2959.88	**2948.71**	NA	NA
61t	3052.02	3048.90	3075.89	**3047.32**	NA	NA
03t	3097.55	**3094.15**	3133.85	3096.52	NA	NA
50t	3067.51	3066.29	3090.02	**3065.50**	NA	NA
52t	3078.96	3077.66	3099.49	**3077.21**	NA	NA
53t	3156.60	3157.434	3157.95	**3155.91**	NA	NA
54t	3243.50	**3243.49**	3268.89	3243.61	NA	NA

Table A.39 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of PM_{10} in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of PM_{10} in 2021						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3051.17	**3045.77**	3077.67	3047.72	NA	NA
10t	**2974.56**	2974.91	2990.51	2975.67	NA	NA
11t	3030.09	**3026.59**	3057.64	3029.42	NA	NA
59t	2819.08	**2810.10**	2862.86	2817.65	NA	NA
61t	2954.92	**2947.44**	2992.15	2954.15	NA	NA
03t	3174.86	**3163.13**	3223.25	3174.78	NA	NA
50t	3071.62	**3069.08**	3097.13	3072.22	NA	NA
52t	3080.46	**3075.01**	3112.41	3079.75	NA	NA
53t	3179.87	**3179.82**	3193.48	3179.89	NA	NA
54t	3147.18	**3143.94**	3184.46	3146.70	NA	NA
BIC of PM_{10} in 2021						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	3070.67	**3065.27**	3097.17	3067.22	NA	NA
10t	**2993.89**	2994.24	3009.84	2995.01	NA	NA
11t	3049.57	**3046.08**	3077.13	3048.91	NA	NA
59t	2838.57	**2829.60**	2882.36	2837.15	NA	NA
61t	2974.33	**2966.86**	3011.57	2973.57	NA	NA
03t	3194.21	**3182.48**	3242.61	3194.13	NA	NA
50t	3091.09	**3088.55**	3116.60	3091.69	NA	NA
52t	3099.96	**3094.51**	3131.91	3099.25	NA	NA
53t	3199.27	**3199.22**	3212.88	3199.30	NA	NA
54t	3166.66	**3163.41**	3203.93	3166.17	NA	NA

Table A.40 The outcomes of the information criteria (AIC and BIC) for 2-mixture distributions of PM_{10} in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

AIC of PM_{10} in 2022						
Station code	AIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2632.68	**2630.32**	2662.53	2630.91	NA	NA
10t	2465.11	**2464.34**	2484.39	2464.68	NA	NA
11t	195.679	195.749	196.540	**195.663**	NA	NA
59t	2681.08	**2672.66**	2735.81	2673.48	NA	NA
61t	876.817	**876.745**	880.598	876.915	NA	NA
03t	911.181	912.692	908.839	**908.371**	NA	NA
50t	1903.87	**1901.72**	1925.91	1902.73	NA	NA
52t	2733.24	**2730.69**	2763.63	2731.07	NA	NA
53t	2448.82	**2448.31**	2460.96	2449.43	NA	NA
54t	3064.26	**3063.61**	3087.56	3063.87	NA	NA
BIC of PM_{10} in 2022						
Station Code	BIC of 2-mixture distributions					
	2GM	2LN	2W	LN-GM	LN-W	GM-W
05t	2651.83	**2649.46**	2681.67	2650.06	NA	NA
10t	2483.72	**2482.94**	2502.99	2483.28	NA	NA
11t	202.158	202.229	203.019	**202.142**	NA	NA
59t	2700.53	**2692.11**	2755.26	2692.93	NA	NA
61t	890.134	**890.062**	893.915	890.232	NA	NA
03t	924.403	925.914	922.061	**921.592**	NA	NA
50t	1921.28	**1919.12**	1943.32	1920.13	NA	NA
52t	2752.54	**2749.99**	2782.93	2750.38	NA	NA
53t	2467.37	**2466.86**	2479.51	2467.98	NA	NA
54t	3083.76	**3083.11**	3107.05	3083.36	NA	NA

A.5 3-Mixture Distributions for $PM_{2.5}$

Table A.41 The outcomes of the information criteria (AIC) for 3-mixture distributions of $PM_{2.5}$ in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2641.17	2637.42	2647.97	**2576.88**	NA
10t	**524.101**	524.559	527.24	524.890	NA
11t	567.160	566.36	567.080	**563.589**	NA
59t	**2089.58**	2170.42	2183.61	2176.55	NA
61t	2403.69	2402.05	2408.04	**2056.76**	NA
03t	**639.029**	639.903	640.460	640.411	NA
50t	**2724.75**	2726.51	2741.09	2729.46	NA
52t	2842.03	2843.99	2859.74	**2841.992**	2844.685
53t	2782.92	2783.35	2789.35	**2782.60**	2788.16
54t	1240.77	**1115.18**	1239.87	1201.804	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2637.37	NA	NA	NA	NA
10t	525.182	NA	NA	NA	NA
11t	566.633	NA	NA	NA	NA
59t	2170.46	NA	NA	NA	NA
61t	2404.19	NA	NA	NA	NA
03t	639.469	NA	NA	NA	NA
50t	2726.18	NA	NA	NA	NA
52t	2842.78	NA	NA	NA	NA
53t	2783.58	NA	NA	NA	NA
54t	1238.43	NA	NA	NA	NA

Table A.42 The outcomes of the information criteria (BIC) for 3-mixture distributions of $PM_{2.5}$ in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2673.11	2668.19	2678.74	**2607.65**	NA
10t	**540.857**	541.096	540.980	542.646	NA
11t	585.215	585.010	585.730	**582.236**	NA
59t	**2119.33**	2203.96	2213.25	2206.18	NA
61t	2433.81	2433.85	2438.16	**2086.89**	NA
03t	**656.505**	658.849	659.110	659.057	NA
50t	**2755.88**	2757.64	2772.22	2756.69	NA
52t	2875.26	2873.04	2890.81	2873.06	**2841.99**
53t	2815.68	2814.28	2820.55	**2813.80**	2815.46
54t	1265.22	**1139.63**	1264.32	1226.25	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2668.15	NA	NA	NA	NA
10t	542.938	NA	NA	NA	NA
11t	585.579	NA	NA	NA	NA
59t	2200.09	NA	NA	NA	NA
61t	2434.31	NA	NA	NA	NA
03t	658.115	NA	NA	NA	NA
50t	2757.31	NA	NA	NA	NA
52t	2873.85	NA	NA	NA	NA
53t	2814.78	NA	NA	NA	NA
54t	1259.82	NA	NA	NA	NA

Table A.43 The outcomes of the information criteria (AIC) for 3-mixture distributions of $PM_{2.5}$ in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	**2745.44**	2748.43	2750.41	2748.64	NA
10t	2645.66	2647.30	2652.97	2645.11	NA
11t	**2674.30**	2674.61	2681.45	2674.44	NA
59t	**2675.59**	2676.87	2677.42	2676.39	NA
61t	**2748.12**	2748.24	2755.60	2754.54	NA
03t	2778.30	2778.39	2781.08	**2777.98**	NA
50t	2727.88	2726.80	2739.11	**2725.80**	NA
52t	2822.58	**2812.43**	2825.83	2822.39	NA
53t	2806.71	**2803.48**	2807.04	2806.19	NA
54t	**2783.31**	2786.57	2810.32	2783.63	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2750.75	NA	NA	NA	NA
10t	**2642.42**	NA	NA	NA	NA
11t	2674.50	NA	NA	NA	NA
59t	2678.31	NA	NA	NA	NA
61t	2748.47	NA	NA	NA	NA
03t	2778.02	NA	NA	NA	NA
50t	2728.82	NA	NA	NA	NA
52t	2823.50	NA	NA	NA	NA
53t	2805.15	NA	NA	NA	NA
54t	2786.47	NA	NA	NA	NA

Table A.44 The outcomes of the information criteria (BIC) for 3-mixture distributions of $PM_{2.5}$ in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2776.62	**2775.71**	2781.59	2775.92	NA
10t	2676.59	2674.37	2683.91	2676.05	NA
11t	**2705.39**	2705.70	2712.54	2705.53	NA
59t	2706.77	2708.52	2708.60	2707.57	**2705.59**
61t	**2779.16**	2779.28	2786.64	2785.59	NA
03t	2809.37	2809.45	2812.15	**2809.05**	NA
50t	2759.03	2757.95	2770.26	**2756.95**	NA
52t	2853.78	**2843.63**	2857.03	2853.59	NA
53t	2837.91	**2834.68**	2807.04	2837.39	NA
54t	**2814.45**	2817.71	2841.45	2814.76	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2781.93	NA	NA	NA	NA
10t	**2673.35**	NA	NA	NA	NA
11t	2705.59	NA	NA	NA	NA
59t	2708.05	NA	NA	NA	NA
61t	2779.51	NA	NA	NA	NA
03t	2809.09	NA	NA	NA	NA
50t	2759.98	NA	NA	NA	NA
52t	2854.70	NA	NA	NA	NA
53t	2836.35	NA	NA	NA	NA
54t	2817.60	NA	NA	NA	NA

Table A.45 The outcomes of the information criteria (AIC) for 3-mixture distributions of $PM_{2.5}$ in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2688.06	2687.10	2693.44	**2686.51**	NA
10t	2561.71	2562.61	2569.05	2562.33	NA
11t	2654.50	**2523.57**	2664.39	2654.50	NA
59t	**2677.23**	2677.35	2683.24	2681.46	NA
61t	2484.48	2486.09	2491.91	2485.12	NA
03t	2733.73	**2717.03**	2722.23	2733.66	NA
50t	2677.44	2677.96	2689.67	**2674.74**	NA
52t	**2696.45**	2697.24	2697.64	2697.27	NA
53t	2779.68	**2777.77**	2778.76	2780.72	NA
54t	2742.21	**2740.11**	2762.43	2742.02	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2686.87	NA	NA	NA	NA
10t	**2555.62**	NA	NA	NA	NA
11t	2655.38	NA	NA	NA	NA
59t	2682.57	NA	NA	NA	NA
61t	**2482.68**	NA	NA	NA	NA
03t	2738.10	NA	NA	NA	NA
50t	2677.29	NA	NA	NA	NA
52t	2697.27	NA	NA	NA	NA
53t	2777.96	NA	NA	NA	NA
54t	2740.98	NA	NA	NA	NA

Table A.46 The outcomes of the information criteria (BIC) for 3-mixture distributions of $PM_{2.5}$ in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2719.29	2718.32	2724.66	**2713.83**	NA
10t	2592.73	2593.64	2600.07	2589.47	NA
11t	2685.68	**2554.75**	2695.57	2685.68	NA
59t	**2708.43**	2708.55	2714.44	2712.66	NA
61t	2515.25	2516.86	2522.68	2515.89	NA
03t	2764.90	**2748.21**	2753.40	2764.83	NA
50t	**2704.66**	2709.07	2720.78	2705.85	NA
52t	**2727.58**	2728.37	2728.78	2727.63	NA
53t	2810.75	**2808.84**	2809.83	2811.79	NA
54t	2773.35	**2771.24**	2793.56	2773.16	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2718.09	NA	NA	NA	NA
10t	**2586.64**	NA	NA	NA	NA
11t	2686.56	NA	NA	NA	NA
59t	2713.77	NA	NA	NA	NA
61t	**2513.45**	NA	NA	NA	NA
03t	2769.28	NA	NA	NA	NA
50t	2708.41	NA	NA	NA	NA
52t	2728.40	NA	NA	NA	NA
53t	2809.03	NA	NA	NA	NA
54t	2772.11	NA	NA	NA	NA

Table A.47 The outcomes of the information criteria (AIC) for 3-mixture distributions of $PM_{2.5}$ in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2754.50	**2749.77**	2757.29	2751.52	NA
10t	2593.94	2593.80	2597.98	2593.64	NA
11t	2692.64	2692.19	2717.98	2692.66	NA
59t	2662.18	**2661.88**	2663.98	2667.71	NA
61t	2287.59	**2287.23**	2308.61	2287.34	NA
03t	2766.11	2767.45	2771.08	2768.06	NA
50t	2678.55	2678.82	2687.27	2689.67	NA
52t	2734.93	**2732.04**	2745.25	2748.34	NA
53t	2820.50	2820.68	2822.08	2822.73	NA
54t	2806.17	2805.77	2816.76	2807.98	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2751.26	NA	NA	NA	NA
10t	**2593.47**	NA	NA	NA	NA
11t	**2691.99**	NA	NA	NA	NA
59t	2662.06	NA	NA	NA	NA
61t	2287.43	NA	NA	NA	NA
03t	**2763.96**	NA	NA	NA	NA
50t	**2678.51**	NA	NA	NA	NA
52t	2748.17	NA	NA	NA	NA
53t	**2818.97**	NA	NA	NA	NA
54t	**2805.25**	NA	NA	NA	NA

Table A.48 The outcomes of the information criteria (BIC) for 3-mixture distributions of $PM_{2.5}$ in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2785.70	**2777.07**	2778.49	2782.72	NA
10t	2624.89	2624.75	2628.94	2624.60	NA
11t	2723.84	**2719.49**	2745.27	2723.86	NA
59t	2693.38	**2693.07**	2695.18	2698.91	NA
61t	2318.65	**2318.32**	2339.67	2318.79	NA
03t	2797.24	2798.67	2802.19	2807.14	NA
50t	2709.67	2709.93	2718.38	2595.93	NA
52t	2766.13	**2763.24**	2776.44	2775.64	NA
53t	2851.54	2851.73	2853.13	**2849.89**	NA
54t	2837.32	2836.93	2847.91	2839.14	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2778.56	NA	NA	NA	NA
10t	**2624.42**	NA	NA	NA	NA
11t	2723.19	NA	NA	NA	NA
59t	2693.26	NA	NA	NA	NA
61t	2318.49	NA	NA	NA	NA
03t	**2796.73**	NA	NA	NA	NA
50t	**2709.62**	NA	NA	NA	NA
52t	2775.39	NA	NA	NA	NA
53t	2850.02	NA	NA	NA	NA
54t	**2836.40**	NA	NA	NA	NA

Table A.49 The outcomes of the information criteria (AIC) for 3-mixture distributions of $PM_{2.5}$ in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2602.06	2602.09	2610.61	2602.50	NA
10t	2410.23	2406.77	2419.91	2406.85	NA
11t	2577.74	**2574.38**	2591.61	2577.06	NA
59t	2478.88	2479.19	2489.99	2483.32	NA
61t	2266.81	2260.42	2308.01	**1307.83**	NA
03t	2611.30	**2608.93**	2622.94	2615.31	NA
50t	2516.99	**2422.37**	2529.36	2516.67	NA
52t	2588.86	2619.61	2596.73	2588.98	NA
53t	2343.91	**2342.46**	2353.91	2344.01	NA
54t	2661.48	**2660.41**	2670.21	2661.47	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	**2601.52**	NA	NA	NA	NA
10t	**2406.76**	NA	NA	NA	NA
11t	2574.75	NA	NA	NA	NA
59t	**2478.03**	NA	NA	NA	NA
61t	2260.69	NA	NA	NA	NA
03t	2612.32	NA	NA	NA	NA
50t	2516.87	NA	NA	NA	NA
52t	**2588.42**	NA	NA	NA	NA
53t	2344.12	NA	NA	NA	NA
54t	2666.24	NA	NA	NA	NA

Table A.50 The outcomes of the information criteria (BIC) for 3-mixture distributions of $PM_{2.5}$ in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2633.21	2633.24	2641.76	2633.66	NA
10t	2440.98	2437.52	2450.66	2437.61	NA
11t	2608.89	**2605.54**	2622.76	2608.21	NA
59t	2510.04	2510.34	2521.15	2510.59	NA
61t	2297.97	2291.57	2339.16	**1338.99**	NA
03t	2642.47	**2640.10**	2654.12	2646.48	NA
50t	2547.67	**2453.37**	2560.36	2547.87	NA
52t	2619.97	2619.61	2627.84	2620.09	NA
53t	2374.05	**2372.61**	2384.06	2374.16	NA
54t	2692.63	2691.56	2701.36	2692.63	**2633.35**
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	**2632.68**	NA	NA	NA	NA
10t	**2437.51**	NA	NA	NA	NA
11t	2605.90	NA	NA	NA	NA
59t	**2505.29**	NA	NA	NA	NA
61t	2291.85	NA	NA	NA	NA
03t	2643.50	NA	NA	NA	NA
50t	2547.87	NA	NA	NA	NA
52t	**2619.53**	NA	NA	NA	NA
53t	2374.27	NA	NA	NA	NA
54t	2693.50	NA	NA	NA	NA

A.6 3-Mixture Distributions for PM_{10}

Table A.51 The outcomes of the information criteria (AIC) for 3-mixture distributions of PM_{10} in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2994.54	**2994.17**	3003.87	2994.24	NA
10t	2458.35	**1375.79**	2461.10	2462.04	NA
11t	**268.835**	268.956	273.970	268.905	NA
59t	2965.72	**2961.47**	2981.97	2964.46	NA
61t	**2652.56**	2653.15	2655.10	2652.59	NA
03t	3219.92	3221.88	3221.69	**3219.61**	NA
50t	**2938.60**	3097.89	3109.32	3097.16	NA
52t	3194.05	3192.79	3199.81	3184.78	NA
53t	3219.34	**3218.67**	3237.36	3219.07	NA
54t	**1370.51**	1425.60	1429.82	1425.50	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2994.45	NA	NA	NA	NA
10t	2459.08	NA	NA	NA	NA
11t	281.098	NA	NA	NA	NA
59t	2961.58	NA	NA	NA	NA
61t	2653.26	NA	NA	NA	NA
03t	3221.63	NA	NA	NA	NA
50t	3098.75	NA	NA	NA	NA
52t	**3184.73**	NA	NA	NA	NA
53t	3218.74	NA	NA	NA	NA
54t	1425.59	NA	NA	NA	NA

Table A.52 The outcomes of the information criteria (BIC) for 3-mixture distributions of PM_{10} in 2018. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3025.47	**3025.10**	3034.80	3025.17	NA
10t	2487.73	**1405.18**	2490.49	2491.43	NA
11t	**281.046**	281.167	286.181	281.116	NA
59t	2996.81	**2992.56**	3013.06	2995.55	NA
61t	**2682.75**	2683.35	2685.30	2682.78	NA
03t	3251.01	3252.97	3252.77	**3250.70**	NA
50t	**2969.67**	3128.96	3140.39	3128.22	NA
52t	3225.12	3223.86	3230.88	3215.85	NA
53t	3250.54	**3249.87**	3268.56	3250.26	NA
54t	**1395.02**	1450.10	1454.32	1450.00	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3025.38	NA	NA	NA	NA
10t	2488.47	NA	NA	NA	NA
11t	285.722	NA	NA	NA	NA
59t	2992.67	NA	NA	NA	NA
61t	2683.46	NA	NA	NA	NA
03t	3252.72	NA	NA	NA	NA
50t	3129.82	NA	NA	NA	NA
52t	**3215.81**	NA	NA	NA	NA
53t	3249.94	NA	NA	NA	NA
54t	1450.09	NA	NA	NA	NA

Table A.53 The outcomes of the information criteria (AIC) for 3-mixture distributions of PM_{10} in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3099.49	**3099.14**	3003.87	3099.19	NA
10t	2978.06	2977.31	2986.17	2977.89	NA
11t	**1441.06**	1445.07	1447.35	1446.52	NA
59t	2914.51	2914.23	2924.80	2914.27	NA
61t	3023.57	3023.63	3038.84	**3023.55**	NA
03t	3208.91	3208.67	3213.10	**3208.35**	NA
50t	3097.02	**3093.78**	3103.83	3096.97	NA
52t	3087.23	**3083.17**	3092.66	3084.58	NA
53t	3173.28	3173.26	3183.36	**3172.81**	NA
54t	3115.80	3115.79	3117.77	**3115.73**	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3099.63	NA	NA	NA	NA
10t	**2977.20**	NA	NA	NA	NA
11t	1444.96	NA	NA	NA	NA
59t	**2914.08**	NA	NA	NA	NA
61t	3023.56	NA	NA	NA	NA
03t	3208.91	NA	NA	NA	NA
50t	3094.11	NA	NA	NA	NA
52t	3086.63	NA	NA	NA	NA
53t	3173.87	NA	NA	NA	NA
54t	3108.74	NA	NA	NA	NA

Table A.54 The outcomes of the information criteria (BIC) for 3-mixture distributions of PM_{10} in 2019. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3130.69	**3130.34**	3134.07	3130.39	NA
10t	3009.11	3008.35	3017.22	3008.94	NA
11t	**1466.15**	1470.16	1472.43	1471.61	NA
59t	2945.71	2945.43	2956.00	2945.47	NA
61t	3054.70	3054.77	3069.97	**3054.68**	NA
03t	3240.02	3239.78	3244.21	**3239.46**	NA
50t	3128.22	**3124.98**	3135.03	3128.17	NA
52t	3118.37	**3114.31**	3123.79	3115.72	NA
53t	3204.46	3204.44	3214.53	**3203.98**	NA
54t	3146.84	3146.83	3148.81	**3146.77**	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3130.83	NA	NA	NA	NA
10t	**3008.24**	NA	NA	NA	NA
11t	1470.04	NA	NA	NA	NA
59t	**2945.28**	NA	NA	NA	NA
61t	3054.69	NA	NA	NA	NA
03t	3240.02	NA	NA	NA	NA
50t	3125.31	NA	NA	NA	NA
52t	3117.76	NA	NA	NA	NA
53t	3205.05	NA	NA	NA	NA
54t	3139.78	NA	NA	NA	NA

Table A.55 The outcomes of the information criteria (AIC) for 3-mixture distributions of PM_{10} in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3050.28	**3049.73**	3058.65	3050.06	NA
10t	**2976.22**	2979.36	2988.42	2977.77	NA
11t	3031.98	**2660.94**	3042.84	3031.07	NA
59t	2812.62	2812.13	2831.92	2812.18	NA
61t	2951.57	2951.36	2966.88	**2951.01**	NA
03t	3161.90	3162.03	3173.05	**3161.72**	NA
50t	3075.63	**3074.31**	3076.74	3076.92	NA
52t	3075.81	**3075.29**	3086.14	3075.94	NA
53t	3185.62	3179.56	3191.11	**3179.38**	NA
54t	3149.04	3147.93	3158.43	**3147.67**	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3049.89	NA	NA	NA	NA
10t	2978.36	NA	NA	NA	NA
11t	3031.30	NA	NA	NA	NA
59t	**2811.92**	NA	NA	NA	NA
61t	2951.75	NA	NA	NA	NA
03t	3166.28	NA	NA	NA	NA
50t	3075.48	NA	NA	NA	NA
52t	3075.92	NA	NA	NA	NA
53t	3184.11	NA	NA	NA	NA
54t	3147.74	NA	NA	NA	NA

Table A.56 The outcomes of the information criteria (BIC) for 3-mixture distributions of PM_{10} in 2020. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3081.48	**3080.93**	3089.85	3081.26	NA
10t	**3007.15**	3010.29	3019.35	3008.70	NA
11t	3063.16	**2692.12**	3074.01	3062.25	NA
59t	2843.82	2843.33	2863.12	2843.38	NA
61t	2982.63	2982.43	2997.95	**2982.07**	NA
03t	3192.87	3192.98	3204.01	**3192.68**	NA
50t	3106.78	3105.46	3107.90	**3102.88**	NA
52t	3107.01	**3106.44**	3117.34	3107.14	NA
53t	3216.82	**3210.61**	3222.16	3216.21	NA
54t	3180.20	3181.06	3189.58	**3177.04**	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3081.09	NA	NA	NA	NA
10t	3009.29	NA	NA	NA	NA
11t	3062.47	NA	NA	NA	NA
59t	**2843.12**	NA	NA	NA	NA
61t	2982.82	NA	NA	NA	NA
03t	3192.86	NA	NA	NA	NA
50t	3106.64	NA	NA	NA	NA
52t	3107.12	NA	NA	NA	NA
53t	3211.27	NA	NA	NA	NA
54t	3178.89	NA	NA	NA	NA

Table A.57 The outcomes of the information criteria (AIC) for 3-mixture distributions of PM_{10} in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3050.25	**3049.88**	3058.65	3055.17	NA
10t	**2978.44**	2978.55	2988.42	2977.77	NA
11t	3032.01	**2782.36**	3042.84	3032.21	NA
59t	2812.66	2812.13	2831.92	2812.60	NA
61t	2951.57	2951.36	2966.88	**2950.97**	NA
03t	3161.90	3162.03	3173.05	**3161.72**	NA
50t	3073.71	**3069.33**	3076.74	3075.62	NA
52t	3075.81	**3075.29**	3086.14	3075.70	NA
53t	3185.77	3179.56	3191.11	**3179.38**	NA
54t	3031.17	3149.91	3158.43	**3145.88**	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	3049.66	NA	NA	NA	NA
10t	2978.98	NA	NA	NA	NA
11t	2841.12	NA	NA	NA	NA
59t	**2812.02**	NA	NA	NA	NA
61t	2951.75	NA	NA	NA	NA
03t	3166.28	NA	NA	NA	NA
50t	3072.79	NA	NA	NA	NA
52t	3075.31	NA	NA	NA	NA
53t	3184.11	NA	NA	NA	NA
54t	3147.74	NA	NA	NA	NA

Table A.58 The outcomes of the information criteria (BIC) for 3-mixture distributions of PM_{10} in 2021. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	3081.45	3081.08	3089.85	3082.47	NA
10t	3009.37	3010.39	3019.35	**3005.61**	NA
11t	3063.16	**2813.53**	3074.01	3063.38	NA
59t	2843.86	2843.33	2863.12	2843.80	NA
61t	2982.84	2982.43	2997.95	**2982.03**	NA
03t	3192.85	3192.98	3204.01	**3192.68**	NA
50t	3104.86	3105.47	3107.90	**3102.88**	NA
52t	3107.01	3107.49	3117.34	**3105.01**	NA
53t	3216.67	3216.86	3222.16	**3210.42**	NA
54t	3062.32	3181.06	3189.58	**3177.04**	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	**3080.86**	NA	NA	NA	NA
10t	3009.92	NA	NA	NA	NA
11t	2872.30	NA	NA	NA	NA
59t	**2843.22**	NA	NA	NA	NA
61t	2982.59	NA	NA	NA	NA
03t	3197.24	NA	NA	NA	NA
50t	3103.94	NA	NA	NA	NA
52t	3106.50	NA	NA	NA	NA
53t	3210.61	NA	NA	NA	NA
54t	3178.89	NA	NA	NA	NA

Table A.59 The outcomes of the information criteria (AIC) for 3-mixture distributions of PM_{10} in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	AIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2635.75	**2634.32**	2648.24	2635.47	NA
10t	2469.11	2469.21	2476.95	**2468.85**	NA
11t	200.065	201.403	203.233	**200.064**	NA
59t	2678.09	2675.54	2714.34	2676.04	NA
61t	872.181	881.538	874.123	**872.175**	NA
03t	909.247	909.156	914.290	**909.006**	NA
50t	1907.71	**1905.43**	1919.59	1905.99	NA
52t	**2734.28**	2734.93	2742.46	2734.61	NA
53t	2451.49	2450.97	2455.62	**2446.43**	NA
54t	3068.47	3069.64	3075.11	3068.84	NA
Station Code	AIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	2552.60	NA	NA	NA	NA
10t	2469.38	NA	NA	NA	NA
11t	200.077	NA	NA	NA	NA
59t	**2675.47**	NA	NA	NA	NA
61t	880.621	NA	NA	NA	NA
03t	909.225	NA	NA	NA	NA
50t	1906.13	NA	NA	NA	NA
52t	2734.36	NA	NA	NA	NA
53t	2451.16	NA	NA	NA	NA
54t	**3067.91**	NA	NA	NA	NA

Table A.60 The outcomes of the information criteria (BIC) for 3-mixture distributions of PM_{10} in 2022. The lowest values are marked in bold. Values that cannot be estimated are indicated by NA.

Station code	BIC of 3-mixture distributions				
	3GM	3LN	3W	GM-GM-LN	GM-GM-W
05t	2666.38	2661.12	2678.87	2666.10	NA
10t	2498.98	2495.97	2506.71	**2495.89**	NA
11t	210.432	211.769	211.600	**203.771**	NA
59t	2712.00	2706.59	2745.45	2708.65	NA
61t	901.160	895.633	895.430	**893.467**	NA
03t	930.402	**930.310**	935.445	931.315	NA
50t	**1932.00**	1935.01	1947.43	1934.13	NA
52t	**2765.17**	2765.49	2773.35	2765.82	NA
53t	2481.17	2480.65	2485.31	**2476.12**	NA
54t	3099.67	3100.84	3106.30	3099.00	NA
Station Code	BIC of 3-mixture distribution				
	GM-LN-LN	GM-LN-W	GM-W-W	LN-LN-W	LN-W-W
05t	**2583.23**	NA	NA	NA	NA
10t	2499.15	NA	NA	NA	NA
11t	210.444	NA	NA	NA	NA
59t	**2706.58**	NA	NA	NA	NA
61t	901.109	NA	NA	NA	NA
03t	930.381	NA	NA	NA	NA
50t	1933.98	NA	NA	NA	NA
52t	2765.24	NA	NA	NA	NA
53t	2480.85	NA	NA	NA	NA
54t	**3098.99**	NA	NA	NA	NA

APPENDIX B

APPLICATION OF R IN STATISTICAL DISTRIBUTIONS OF DATA ANALYSIS

This appendix presents some R code using in this thesis.

```
# For analysis of the PM2.5 and PM10, there are steps to do the same.
# Click link and download data :
# At https://pcd.gdcatalog.go.th/
# Save data to appropriate folder .
# Step 1: Import data from the stored f il e using the ''utils library .
Data <- read.csv( 'File Address . 'csv )

# Select variables PM2.5 and PM10 concentrations from 10 stations in Bangkok
# for analysis from 2018 to 2022;
Station_{1}= 05t
Station_{2}= 10t
Station_{3}= 11t
Station_{4}= 59t
Station_{5}= 61t
Station_{6}= 03t
Station_{7}= 50t
Station_{8}= 52t
Station_{9}= 53t
Station_{10}= 54t

# Step 2.1: Utilizing the 'fitdistrplus' library for fitting data with a no mixture distribution:
# the gamma distribution, the lognormal distribution, and the Weibull distribution.
# and evaluate information criteria.
fitgamma <- fitdist(data, ''gamma", method = "MLE")
fitgamma <- fitdist(data, ''lognormal", method = "MLE")
fitgamma <- fitdist(data, ''Weibull", method = "MLE")

# Step 2.2: Utilizing the 'stats' library for test the goodness-of-fit.
# Example case of gamma distribution.
# In the case of lognormal distribution, and Weibull distribution, the same is done.
KS <- ks.test(data, pgamma, shape, scale )
CM <- cvm.test(data, pgamma, shape, scale )
AD <- ad.test(data, pgamma, shape, scale )

# Step 3: Utilizing the 'ltmix' library for fitting data
# with the mixture distribution and evaluate information criteria.
2mix <- ltmm(data, G = 2, distributions = c('gamma', 'lognormal', 'Weibull'), method = "MLE")
3mix <- ltmm(data, G = 3, distributions = c('gamma', 'lognormal', 'Weibull'), method = "MLE")

# Step 4.1: Utilizing the 'extRemes' library for fitting data
# with the generalized extreme value distribution (GEV) and evaluate information criteria.
# In the generalized extreme value distribution, the data select the highest value of each month.
gevfit <- fevd(data, type = ''GEV", method = "MLE", period.basis = "months" )

# Step 4.2: Utilizing the 'gnFit' library for evaluate goodness-of-fit tests (CM and AD)
```

```
# of the generalized extreme value distribution.
parametergev <- gevfit$mle
goodnessoffitgev <- gnfit(data, "gev", pr = parametergev)

# Step 4.3: Utilizing the 'extRemes' library for estimate return levels of GEV.
# In GEV, the data select the highest value of each month.
# The return periods used are 24 (2 years), 60 (5 years), 120 (10 years), and 180 (15 years).
returnlevelgev <- ci(gevfit , return.period = c(24, 60, 120, 180))

# Step 5.1: Utilizing the 'extRemes' library for fitting data
# with the generalized Pareto distribution (GPD) and evaluate information criteria.
# In GPD, the data select the values that exceeded these threshold
# and used the 'stats' library for examined three thresholds derived
# from the quantiles: 90%, 95%, and 99%.
threshold90 <- quantile(data, 0.9 )
gpdfit90 <- fevd(data, threshold = threshold90, type = ''GP" )
threshold95 <- quantile(data, 0.95 )
gpdfit95 <- fevd(data, threshold = threshold95, type = ''GP" )
threshold99 <- quantile(data, 0.99)
gpdfit99 <- fevd(data, threshold = threshold99, type = ''GP" )

# Step 5.2: Utilizing the 'gnFit' library for evaluate goodness-of-fit tests (CM and AD) of GPD.
parameter90 <- gpdfit90$mle
goodnessoffit90 <- gnfit(data, "gpd", pr = parameter90, threshold = threshold90)
parameter95 <- gpdfit95$mle
goodnessoffit95 <- gnfit(data, "gpd", pr = parameter95, threshold = threshold95)
parameter99 <- gpdfit99$mle
goodnessoffit99 <- gnfit(data, "gpd", pr = parameter99, threshold = threshold99)

# Step 5.3: Utilizing the 'extRemes' library for estimate return levels of the GPD.
# In GPD, the data select the values that exceeded these threshold.
# The return periods used are 2 years, 5 years, 10 years, and 15 years.
returnlevel90 <- ci(gpdfit90, return.period = c(2, 5, 10, 15))
returnlevel95 <- ci(gpdfit95, return.period = c(2, 5, 10, 15))
returnlevel99 <- ci(gpdfit99, return.period = c(2, 5, 10, 15))
```

CURRICULUM VITAE

NAME : Tanatip Hanpayak **GENDER :** Female

EDUCATION BACKGROUND:

- Bachelor of Sciences (Mathematical), Suranaree University of Technology, Thailand, 2021

SCHOLARSHIP:

- Outstanding academic performers for graduate studies scholarship of Suranaree University of Technology

CONFERENCE:

- Hanpayak, T., and Areerak, T. (2023). A Study on the Statistical Distribution of $PM_{2.5}$ and PM_{10} in Bangkok., *The 6th CRU-National Conference in Science and Technology : NCST 6th 2023 (online conference)*, Chandrakasem Rajabhat University, Bangkok, 14 June 2023, F51-F62.

EXPERIENCE:

- Teaching assistant in Suranaree University of Technology, Advanced Calculus (Thai course), Analytical Calculus I (Thai course), Calculus II (Thai course), Calculus III (Thai course), Differential Equations I (Thai course)