symmetry

MDPI

Article

Statistical Estimation of Common Percentile in Birnbaum–Saunders Distributions: Insights from $PM_{2.5}$ Data in Thailand

Warisa Thangjai [1], Sa-Aat Niwitpong [2,*], Suparat Niwitpong [2] and Rattana Prommai [3]

[1] Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok 10240, Thailand; wthangjai@rumail.ru.ac.th
[2] Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand; suparat.n@sci.kmutnb.ac.th
[3] Department of Data Science and Application, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand; rattana.pr@up.ac.th
* Correspondence: sa-aat.n@sci.kmutnb.ac.th

Abstract

This study develops approaches for estimating the common percentile of Birnbaum– Saunders (BS) distributions and applies them to daily $PM_{2.5}$ concentration data from six monitoring stations in Chiang Mai Province, Thailand. Percentiles provide a robust representation of typical pollutant exposure, being less sensitive to outliers and suitable for skewed environmental data. Estimating the same percentile across multiple monitoring sites offers a standardized metric for regional air quality assessment, enabling meaningful comparisons and informing evidence-based environmental policy. Four statistical approaches—Generalized Confidence Interval (GCI), bootstrap, Bayesian, and Highest Posterior Density (HPD)—were employed to construct confidence intervals (CIs) for the common percentile. Simulation studies evaluated the methods in terms of average length (AL) and coverage probability (CP), showing that the GCI approach offers the best balance between precision and reliability. Application to real $PM_{2.5}$ data confirmed that the BS distribution appropriately models pollutant concentrations and that the common percentile provides a meaningful measure for environmental assessment. These findings highlight the GCI method as a robust tool for constructing CIs in environmental data analysis.

Keywords: Birnbaum–Saunders model; common percentile estimation; confidence intervals; $PM_{2.5}$ exposure; environmental statistics

Check for updates

Academic Editor: Qing-Wen Wang

Received: 24 November 2025
Revised: 2 January 2026
Accepted: 4 January 2026
Published: 6 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Percentiles based on the Birnbaum–Saunders (BS) distribution offer notable benefits for statistical inference, particularly for positively skewed and nonnegative data commonly encountered in lifetime analysis, reliability engineering, and environmental studies. Relative to moment-based summaries, BS percentiles are more resistant to the influence of extreme observations and provide meaningful interpretation at specified risk or exposure levels. One key advantage of the BS framework is that its percentile function is available in closed form, enabling efficient computation, facilitating bootstrap and resampling-based inference, and allowing confidence intervals to be obtained without numerical inversion. Moreover, BS percentiles effectively describe upper-tail behavior arising from fatigue and cumulative damage mechanisms, making them especially appropriate for inference on

high-order quantiles that are often of primary practical interest. Overall, these features establish BS percentiles as a flexible and computationally efficient alternative to mean-based or purely empirical quantile methods when the focus lies on tail-related characteristics.

Extensive research has addressed statistical inference for percentiles. Marshall and Walsh [1] were among the first to compare percentage points from two continuous populations. Harrell and Davis [2] introduced a quantile estimator, and Kaigh and Lachenbruch [3] later offered a different estimation method. Albers and Löhnberg [4] proposed distribution-free confidence intervals (CIs) for quantile differences in biomedical studies. Cox and Jaber [5] explored tests for the equality of percentiles in normal distributions. Chang and Tang [6] constructed CIs for BS percentiles, while Padgett and Tomlinson [7] derived lower confidence bounds for percentiles from Weibull and BS distributions. Guo and Krishnamoorthy [8] developed Generalized Confidence Intervals (GCIs) and *p*-values to compare quantiles of two normal populations. Huang and Johnson [9] focused on inference regarding the ratio of two normal percentiles. More recent work includes Navruz and Özdemir [10], who introduced a bootstrap approach for comparing percentiles between independent groups, and Hasan and Krishnamoorthy [11], who constructed CIs for the mean and upper percentiles of nonnegative log-normal data. Abdollahnezhad and Jafari [12] additionally proposed a small-sample adjustment to the likelihood ratio test for assessing quantile equality across several normal populations.

Fine particulate matter ($PM_{2.5}$) is a major component of air pollution and remains a serious challenge to both environmental sustainability and human health. High concentrations of $PM_{2.5}$ are commonly observed in various parts of the world, including northern Thailand, where levels often surpass international safety standards and pose significant public health risks. Tackling this issue effectively depends on reliable environmental monitoring and the use of sophisticated data analysis techniques to uncover pollution trends and guide policy decisions. $PM_{2.5}$ consists of very fine airborne particles with diameters under 2.5 micrometers. Because of their tiny size, they can penetrate deep into the lungs and even enter the bloodstream, leading to significant health hazards. Prolonged or high-level exposure has been associated with respiratory and cardiovascular diseases, as well as increased mortality. Key sources of $PM_{2.5}$ include emissions from vehicles, industrial activities, biomass burning, and natural events such as wildfires and dust storms. Urban centers and rapidly developing regions are particularly at risk due to high population density and concentrated human activity. In northern Thailand, $PM_{2.5}$ concentrations regularly exceed the limits set by the World Health Organization (WHO), intensifying concerns about environmental health and long-term sustainability. Consequently, monitoring $PM_{2.5}$ levels and understanding their patterns is essential for evaluating environmental and health effects. Sophisticated statistical analysis is essential for detecting patterns, predicting pollution spikes, and developing effective mitigation strategies. Because $PM_{2.5}$ data are often skewed due to sporadic but severe pollution events, traditional summary measures like the mean may not accurately capture typical exposure levels. In contrast, percentiles offer a more robust and representative measure, making them an increasingly important tool in evidence-based air quality management and policy formulation. Daily $PM_{2.5}$ concentrations from six monitoring stations in Chiang Mai Province, Thailand, were analyzed due to their non-negative and positively skewed distribution, characteristics well captured by the BS model. The BS distribution is particularly suitable for environmental pollutants like $PM_{2.5}$, which can exhibit extreme values during episodic pollution events. Chiang Mai's pronounced seasonal air quality variability, especially during the biomass-burning period, provides an ideal context to demonstrate the robustness of the BS distribution and the estimation of common percentiles across multiple monitoring sites.

This study emphasizes the application of the BS distribution, a flexible statistical model specifically suited for positively skewed data, to analyze $PM_{2.5}$ concentrations. The suitability of this distribution lies in its ability to accurately capture the asymmetry in pollutant levels arising from irregular but severe pollution episodes, such as seasonal biomass burning or industrial emissions. By estimating a common percentile across multiple monitoring stations, the study introduces a unified and consistent metric for regional air quality assessment (Chang and Tang [6] and Padgett and Tomlinson [7]). This approach not only simplifies comparative analysis across different locations but also enhances the reliability of environmental evaluations. A common percentile serves as a central measure of air pollution levels, facilitating data-driven decisions for public health management and environmental policy formulation. Furthermore, using percentiles within the BS modeling framework helps quantify typical exposure levels more effectively than traditional methods (Harrell and Davis [2] and Guo and Krishnamoorthy [8]). This is particularly valuable for stakeholders and decision-makers who must prioritize resources, design mitigation strategies, and set regulatory standards based on realistic and population-relevant exposure metrics. In summary, percentile-based analysis, supported by the BS distribution, provides a meaningful and statistically sound approach for assessing $PM_{2.5}$ pollution. It enables a comprehensive and harmonized understanding of air quality across regions, ultimately contributing to more effective environmental monitoring and public health protection. There has been substantial scholarly attention devoted to statistical inference for the parameters of the BS distribution. Guo et al. [13] investigated inference procedures for estimating a common mean across multiple BS populations. Shakil et al. [14] conducted a comprehensive study of inference methods for the three-parameter BS distribution, highlighting its statistical properties, theoretical characterizations, and practical applications. More recently, Thangjai et al. [15] developed a GCI approach for comparing percentiles of BS distributions, with an application to $PM_{2.5}$ concentrations in Thailand. Earlier work by Leiva et al. [16] extended the BS model to a generalized form suitable for modeling pollution data, including both positive and negative moments. Wang [17] further contributed by proposing generalized interval estimation techniques for the BS distribution. Additionally, Niu et al. [18] compared two hypothesis testing methods for BS parameters—one employing the exact generalized *p*-value approach and the other relying on the Delta method—highlighting their relative performance.

A common percentile represents the same relative position across multiple distributions. When data are collected from different monitoring stations, each site may have its own distribution, but the common percentile approach assumes that a fixed percentile level (e.g., the 75th) is shared across all locations, even if the corresponding concentration values differ. This provides a unified and interpretable measure of typical exposure that supports meaningful comparisons while accounting for local variability. Evaluating uncertainty around common percentiles is essential for reliable environmental assessments and informed policy decisions. This study explores four statistical approaches for constructing CIs: GCI, bootstrap, Bayesian, and Highest Posterior Density (HPD). These methods are all implemented using R software (version 2024.12.0+467), a popular tool for statistical analysis and data processing. Although frequently used to estimate parameters such as means and variances, these techniques are less commonly applied to percentiles, despite the critical role percentiles play in environmental studies where data distributions are often skewed by occasional extreme pollution events. The versatility of R software (version 2024.12.0+467) facilitates the effective application of these techniques: bootstrap methods rely on resampling to empirically build CIs; Bayesian and HPD approaches use Markov Chain Monte Carlo (MCMC) simulations to approximate posterior distributions; and GCI approach involves programming generalized pivotal quantities (GPQs) for interval esti-

mation. By comparing these R-based methods for estimating CIs of percentiles, the study offers valuable insights into their comparative benefits and practical uses. This knowledge aids researchers and policymakers in choosing robust statistical tools that enhance the precision of air quality assessments, improve uncertainty estimation, and support better environmental decision-making. The study evaluates the performance of each method through simulation experiments, considering criteria such as average length (AL) and coverage probability (CP). Subsequently, the methods are applied to actual daily $PM_{2.5}$ data collected from six monitoring stations in Chiang Mai Province, Thailand. The findings underscore the practical effectiveness of each approach and contribute to the advancement of reliable techniques for environmental data analysis and policy development.

2. Common Percentile

Let X_{ij} be a random variable that follows a BS distributions with shape parameter α_i and scale parameter β_i, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. According to Thangjai et al. [15], the percentile of X_{ij} is

$$\theta_i = \frac{\beta_i}{4}\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2, \tag{1}$$

where $z_p = \Phi^{-1}(p)$ is the p-th quantile of the standard normal distribution.

The estimator of β_i is defined as

$$\hat{\beta}_i = \begin{cases} \max(\hat{\beta}_{1i}, \hat{\beta}_{2i}); & T_i \leqslant 0 \\ \min(\hat{\beta}_{1i}, \hat{\beta}_{2i}); & T_i > 0 \end{cases}, \tag{2}$$

where $\hat{\beta}_{1i}$ and $\hat{\beta}_{2i}$ are two solutions of $\hat{\beta}_i$. The value of $\hat{\beta}_i$ can be obtained by solving

$$A_i\left(\hat{\beta}_i\right)^2 - 2B_i\hat{\beta}_i + C_i = 0, \tag{3}$$

where $A_i = (n_i - 1)I_i^2 - \frac{1}{n_i}L_iT_i^2$, $B_i = (n_i - 1)I_iJ_i - (1 - I_iJ_i)T_i^2$, $C_i = (n_i - 1)I_i^2 - \frac{1}{n_i}K_iT_i^2$, $I_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\sqrt{X_{ij}}$, $J_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{\sqrt{X_{ij}}}$, $K_i = \sum_{j=1}^{n_i}\left(\sqrt{X_{ij}} - I_i\right)^2$, $L_i = \sum_{j=1}^{n_i}\left(\frac{1}{\sqrt{X_{ij}}} - J_i\right)^2$ and $T_i \sim t(n_i - 1)$ is t-distribution with $(n_i - 1)$ degrees of freedom.

The variance of $\hat{\beta}_i$ is

$$Var\left(\hat{\beta}_i\right) = \frac{\alpha_i^2\beta_i^2}{n_i}\left(\frac{1 + \frac{3}{4}\alpha_i^2}{\left(1 + \frac{1}{2}\alpha_i^2\right)^2}\right). \tag{4}$$

The estimator of α_i is defined as

$$\hat{\alpha}_i = \sqrt{\frac{\hat{\beta}_i^2\sum_{j=1}^{n_i}\frac{1}{X_{ij}} - 2n_i\hat{\beta}_i + \sum_{j=1}^{n_i}X_{ij}}{\hat{\beta}_iV_i}}, \tag{5}$$

where $V_i \sim \chi^2(n_i)$ is the chi-squared distribution with n_i degrees of freedom.

The variance of $\hat{\alpha}_i$ is

$$Var(\hat{\alpha}_i) = \frac{\alpha_i^2}{2n_i}. \tag{6}$$

The estimator of θ_i is

$$\hat{\theta}_i = \frac{\hat{\beta}_i}{4}\left(\hat{\alpha}_i z_p + \sqrt{\hat{\alpha}_i^2 z_p^2 + 4}\right)^2. \tag{7}$$

From Appendix A, the variance of $\hat{\theta}_i$ is

$$\begin{aligned} Var(\hat{\theta}_i) &= \frac{\alpha_i^2 \beta_i^2}{8n_i}\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2 \\ &* \left(\left(z_p + \frac{\alpha_i z_p^2}{\sqrt{\alpha_i^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\alpha_i^2}{2\left(2 + \alpha_i^2\right)^2}\right)\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2\right). \end{aligned} \tag{8}$$

The estimate of the variance of $\hat{\theta}_i$ is

$$\begin{aligned} \hat{Var}(\hat{\theta}_i) &\approx \frac{\hat{\alpha}_i^2 \hat{\beta}_i^2}{8n_i}\left(\hat{\alpha}_i z_p + \sqrt{\hat{\alpha}_i^2 z_p^2 + 4}\right)^2 \\ &* \left(\left(z_p + \frac{\hat{\alpha}_i z_p^2}{\sqrt{\hat{\alpha}_i^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\hat{\alpha}_i^2}{2\left(2 + \hat{\alpha}_i^2\right)^2}\right)\left(\hat{\alpha}_i z_p + \sqrt{\hat{\alpha}_i^2 z_p^2 + 4}\right)^2\right). \end{aligned} \tag{9}$$

The common percentile $\hat{\theta}$ is calculated as the weighted average of $\hat{\theta}_i$ across the k individual sample. It is defined as

$$\hat{\theta} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{\hat{Var}(\hat{\theta}_i)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\hat{Var}(\hat{\theta}_i)}. \tag{10}$$

2.1. Generalized Confidence Interval Approach

A GCI is a statistical method designed to create CIs in complex or non-standard situations where conventional approaches either fail or are unsuitable. This modern and versatile technique is especially valuable when traditional CI methods fall short. GCIs utilize simulations and the concept of GPQs to produce reliable intervals in challenging inference problems. Traditional CIs typically depend on well-understood sampling distributions, such as the sample mean following a normal distribution, as justified by the Central Limit Theorem. However, in many practical situations, the sampling distribution might be unknown or too complicated to derive, or the estimator may involve nuisance parameters, which are variables that are not of primary interest but still influence the distribution. Additionally, the parameter being estimated may be a complex function of multiple random variables. To overcome these difficulties, the GCI method constructs a GPQ, which is a random variable that depends on both the observed data and a simulated random variable with a known distribution. When the sample data is plugged in, the GPQ becomes free of any unknown parameters. By simulating the distribution of the GPQ and using its empirical percentiles, one can determine confidence bounds corresponding to a desired confidence level.

The GPQ for θ_i is constructed using the GPQs of β_i and α_i. According to Sun [19], the GPQ of β_i is

$$R_{\beta_i} = \begin{cases} \max(\beta_{1i}, \beta_{2i}); & T_i \leqslant 0 \\ \min(\beta_{1i}, \beta_{2i}); & T_i > 0 \end{cases}, \tag{11}$$

where β_{1i} and β_{2i} are two solutions for β_i.

According to Wang [17], the GPQ of α_i is

$$R_{\alpha_i} = \sqrt{\frac{R_{\beta_i}^2 \sum_{j=1}^{n_i} \frac{1}{X_{ij}} - 2n_i R_{\beta_i} + \sum_{j=1}^{n_i} X_{ij}}{R_{\beta_i} V_i}}. \tag{12}$$

The GPQ for θ_i is

$$R_{\theta_i} = \frac{R_{\beta_i}}{4}\left(R_{\alpha_i} z_p + \sqrt{R_{\alpha_i}^2 z_p^2 + 4}\right)^2. \tag{13}$$

The estimate of variance of R_{θ_i} is

$$\begin{aligned} R_{Var(\hat{\theta}_i)} \approx & \frac{R_{\alpha_i}^2 R_{\beta_i}^2}{8n_i}\left(R_{\alpha_i} z_p + \sqrt{R_{\alpha_i}^2 z_p^2 + 4}\right)^2 \\ * & \left(\left(z_p + \frac{R_{\alpha_i} z_p^2}{\sqrt{R_{\alpha_i}^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3R_{\alpha_i}^2}{2\left(2 + R_{\alpha_i}^2\right)^2}\right)\left(R_{\alpha_i} z_p + \sqrt{R_{\alpha_i}^2 z_p^2 + 4}\right)^2\right), \end{aligned} \tag{14}$$

where $z_p = \Phi^{-1}(p)$ is the p-th quantile of the standard normal distribution and R_{β_i} is defined in Equation (11), and R_{α_i} is defined in Equation (12).

The GPQ for the common percentile is

$$R_\theta = \sum_{i=1}^{k} \frac{R_{\theta_i}}{R_{Var(\hat{\theta}_i)}} \Bigg/ \sum_{i=1}^{k} \frac{1}{R_{Var(\hat{\theta}_i)}}. \tag{15}$$

Therefore, the $100(1-\gamma)\%$ two-sided GCI is

$$CI_{GCI} = [L_{GCI}, U_{GCI}] = [R_\theta(\gamma/2), R_\theta(1-\gamma/2)], \tag{16}$$

where $R_\theta(\gamma/2)$ and $R_\theta(1-\gamma/2)$ denote the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th percentiles of R_θ, respectively. Algorithm 1 was used to construct the GCI.

Algorithm 1 GCI

Step 1: Compute A_i, B_i, and C_i
Step 2: At the m step
(a) Simulate $T_i \sim t(n_i - 1)$ and compute R_{β_i}. If $R_{\beta_i} < 0$, resimulate $T_i \sim t(n_i - 1)$
(b) Simulate $V_i \sim \chi^2(n_i)$ and compute R_{α_i}, R_{θ_i}, $R_{Var(\hat{\theta}_i)}$, and R_θ
Step 3: Repeat step 2 a total of M times to obtain an array of R_θ values
Step 4: Compute L_{GCI} and U_{GCI}

2.2. Bootstrap Approach

A bootstrap CI is a statistical technique used to estimate the range in which a population parameter is likely to lie, based solely on sample data. It is particularly useful when the underlying distribution of the data is unknown or too complex to model analytically. This method involves repeatedly resampling the original dataset with replacement. It is also effective for estimating CIs for complex statistics where traditional methods may be inadequate.

Let x_{ij} be a random sample drawn from BS distributions with parameters α_i and β_i, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. Let x_{ij}^* be a bootstrap sample drawn from BS distributions with parameters $\hat{\alpha}_i$ and $\hat{\beta}_i$, where $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i$ and $\hat{\alpha}_i$ and $\hat{\beta}_i$

are the maximum likelihood estimators of α_i and β_i, respectively. Then, the $\hat{\alpha}_i^*$ and $\hat{\beta}_i^*$ are obtained using B bootstrap samples. The respective correct estimates for $\hat{\alpha}_i^*$ and $\hat{\beta}_i^*$ are

$$\widetilde{\alpha}_{i(r)} = \hat{\alpha}_{i(r)}^* - 2\hat{b}(\hat{\alpha}_i, \alpha_i) \tag{17}$$

and

$$\widetilde{\beta}_{i(r)} = \hat{\beta}_{i(r)}^* - 2\hat{b}(\hat{\beta}_i, \beta_i), \tag{18}$$

where $\hat{b}(\hat{\alpha}_i, \alpha) = \frac{1}{B}\sum_{r=1}^{B}\hat{\alpha}_{i(r)}^* - \hat{\alpha}_i$, $\hat{b}(\hat{\beta}_i, \beta) = \frac{1}{B}\sum_{r=1}^{B}\hat{\beta}_{i(r)}^* - \hat{\beta}_i$, and $r = 1, 2, \ldots, B$.

The bootstrap estimator of θ_i is

$$\hat{\theta}_{i(r)} = \frac{\widetilde{\beta}_{i(r)}}{4}\left(\widetilde{\alpha}_{i(r)}z_p + \sqrt{\widetilde{\alpha}_{i(r)}^2 z_p^2 + 4}\right)^2. \tag{19}$$

The bootstrap estimator of the variance of $\hat{\theta}_{i(r)}$ is

$$\begin{aligned} \hat{Var}\left(\hat{\theta}_{i(r)}\right) &\approx \frac{\widetilde{\alpha}_{i(r)}^2\widetilde{\beta}_{i(r)}^2}{8n_i}\left(\widetilde{\alpha}_{i(r)}z_p + \sqrt{\widetilde{\alpha}_{i(r)}^2 z_p^2 + 4}\right)^2 \\ &* \left(\left(z_p + \frac{\widetilde{\alpha}_{i(r)}z_p^2}{\sqrt{\widetilde{\alpha}_{i(r)}^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\widetilde{\alpha}_{i(r)}^2}{2\left(2 + \widetilde{\alpha}_{i(r)}^2\right)^2}\right)\left(\widetilde{\alpha}_{i(r)}z_p + \sqrt{\widetilde{\alpha}_{i(r)}^2 z_p^2 + 4}\right)^2\right). \end{aligned} \tag{20}$$

The bootstrap estimator of the common percentile is

$$\hat{\theta}_r = \sum_{i=1}^{k}\frac{\hat{\theta}_{i(r)}}{\hat{Var}\left(\hat{\theta}_{i(r)}\right)} \Bigg/ \sum_{i=1}^{k}\frac{1}{\hat{Var}\left(\hat{\theta}_{i(r)}\right)}. \tag{21}$$

Therefore, the $100(1-\gamma)\%$ two-sided bootstrap CI is

$$CI_{BCI} = [L_{BCI}, U_{BCI}] = \left[\hat{\theta}_r(\gamma/2), \hat{\theta}_r(1-\gamma/2)\right], \tag{22}$$

where $\hat{\theta}_r(\gamma/2)$ and $\hat{\theta}_r(1-\gamma/2)$ denote the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th percentiles of $\hat{\theta}_r$, respectively. Algorithm 2 was used to construct the bootstrap CI.

Algorithm 2 Bootstrap CI

Step 1: Specify the values of x_{ij}
Step 2: At the b step
(a) Generate x_{ij}^*
(b) Compute $\widetilde{\alpha}_{i(r)}$, $\widetilde{\beta}_{i(r)}$, $\hat{\theta}_{i(r)}$, $\hat{Var}\left(\hat{\theta}_{i(r)}\right)$, and $\hat{\theta}_r$
Step 3: Repeat step 2 a total of B times to obtain an array of $\hat{\theta}_r$ values
Step 4: Compute L_{BCI} and U_{BCI}

2.3. Bayesian Approach

The Bayesian approach is a statistical inference method that views probability as a reflection of belief or confidence, rather than relying solely on long-run frequencies. It is based on Bayes' Theorem, which provides a way to revise the likelihood of a hypothesis as new data or evidence becomes available. A prior reflects existing knowledge or assumptions about the parameters or model before any data is observed. Priors may be informative, drawing from previous research, or non-informative, implying little to no prior knowledge. The likelihood expresses the probability of the observed data given various parameter values. The posterior represents the updated belief after incorporating the data, which is the central aim of Bayesian inference: to determine the posterior distribution of the model parameters or hypotheses. Key benefits of the Bayesian approach include its

ability to incorporate prior information, its production of full probability distributions, and its flexibility in complex modeling scenarios.

The Bayesian approach offers a systematic way to include prior knowledge and adjust beliefs as new data becomes available. This is particularly useful in complex models where parameters are supported by theoretical insights or empirical evidence. Within this framework, the posterior distribution is derived using Bayes' Theorem, allowing the model to be updated progressively as more information is gathered.

In the study by Wang et al. [20], inverse gamma distributions were used as prior distributions for the parameters β_i and α_i^2. Suppose that β_i follows an inverse gamma distribution with parameters a_{i1} and b_{i1}, denoted as $IG(\beta_i|a_{i1}, b_{i1})$. Similarly, suppose that α_i^2 follows an inverse gamma distribution with parameters a_{i2} and b_{i2}, denoted as $IG(\alpha_i^2|a_{i2}, b_{i2})$. The marginal distribution of β_i is

$$\begin{aligned} p(\beta_i|x_i) &\propto \beta_i^{-(n_i+a_{i1}+1)} \exp\left(-\frac{b_{i1}}{\beta_i}\right) \\ &* \prod_{j=1}^{n_i}\left(\left(\frac{\beta_i}{x_{ij}}\right)^{\frac{1}{2}} + \left(\frac{\beta_i}{x_{ij}}\right)^{\frac{3}{2}}\right)\left(\sum_{j=1}^{n_i}\frac{1}{2}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right) + b_{i2}\right)^{-\frac{n_i+1}{2}-a_{i2}}. \end{aligned} \quad (23)$$

The conditional posterior distribution of α_i^2 given β_i is expressed as

$$p(\alpha_i^2|x_i, \beta_i) \propto IG\left(\frac{n_i}{2} + a_{i2}, \frac{1}{2}\sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right) + b_{i2}\right). \quad (24)$$

The samples shown in Equations (23) and (24) are generated using MCMC methods, a common approach in Bayesian analysis for estimating complex posterior distributions. MCMC algorithms produce a sequence of dependent samples that, under appropriate conditions, converge to the desired posterior distribution. These methods are especially useful when analytical solutions are difficult or impossible to obtain, such as in models with high dimensionality or non-standard probability structures.

Wang et al. [20] utilized the generalized ratio-of-uniforms technique to generate posterior samples for the parameter β_i. This technique was originally introduced by Wakefield et al. [21]. This technique relies on constructing a pair of random variables (u_i, v_i), each drawn from a uniform distribution over a specifically defined region. That is,

$$A(r_i) = \left\{(u_i, v_i) : 0 < u_i \leqslant \left[p\left(\frac{v_i}{u_i^{r_i}}\middle| x_i\right)\right]^{\frac{1}{r_i+1}}\right\}, r_i \geqslant 0, \quad (25)$$

where r_i is a constant and $p(\cdot|x_i)$ is defined according to Equation (23). As a result, the variable $\beta_i = \frac{v_i}{u_i^{r_i}}$ has a probability density proportional to $\frac{p(\beta_i|x_i)}{\int p(\beta_i|x_i)d\beta_i}$. To generate random samples from the region $A(r_i)$, the variables (u_i, v_i) are drawn uniformly from a one-dimensional bounded rectangle that encloses the region $[0, a(r_i)] \times [b^-(r_i), b^+(r_i)]$, where $a(r_i) = \sup_{\beta_i>0}\left\{[p(\beta_i|x_i)]^{\frac{1}{r_i+1}}\right\}$, $b^-(r) = \inf_{\beta_i>0}\left\{\beta_i[p(\beta_i|x_i)]^{\frac{r_i}{r_i+1}}\right\}$, and $b^+(r_i) = \sup_{\beta_i>0}\left\{\beta_i[p(\beta_i|x_i)]^{\frac{r_i}{r_i+1}}\right\}$.

Wang et al. [20] assumed that $a(r_i)$ and $b^+(r_i)$ are finite and that the condition $b^-(r_i) = 0$ holds. A proposed variate $\beta_i = \frac{v_i}{u_i^{r_i}}$ is accepted if $u_i \leqslant [p(\beta_i|x_i)]^{\frac{1}{r_i+1}}$ it satisfies this condition; otherwise, the sampling procedure is repeated. Posterior samples for the parameter α_i^2 are obtained using the LearnBayes package in the R programming

environment. Consequently, the square root of the sampled values α_i^2 corresponds to the posterior samples of the target parameter α_i.

The posterior distribution of θ_i is

$$\theta_{i.Baye} = \frac{\beta_i}{4}\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2. \tag{26}$$

The estimator of the variance of $\theta_{i.Baye}$ is

$$\begin{aligned} \hat{Var}\left(\theta_{i.Baye}\right) \approx & \frac{\alpha_i^2 \beta_i^2}{8n_i}\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2 \\ & * \left(\left(z_p + \frac{\alpha_i z_p^2}{\sqrt{\alpha_i^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\alpha_i^2}{2\left(2 + \alpha_i^2\right)^2}\right)\left(\alpha_i z_p + \sqrt{\alpha_i^2 z_p^2 + 4}\right)^2\right). \end{aligned} \tag{27}$$

The Bayesian estimator of the common percentile is

$$\theta_{Baye} = \sum_{i=1}^{k} \frac{\theta_{i.Baye}}{\hat{Var}\left(\theta_{i.Baye}\right)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\hat{Var}\left(\theta_{i.Baye}\right)}. \tag{28}$$

Therefore, the $100(1-\gamma)\%$ two-sided Bayesian credible interval is

$$CI_{Baye} = \left[L_{Baye}, U_{Baye}\right] = \left[\theta_{Baye}(\gamma/2), \theta_{Baye}(1-\gamma/2)\right], \tag{29}$$

where $\theta_{Baye}(\gamma/2)$ and $\theta_{Baye}(1-\gamma/2)$ denote the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th percentiles of θ_{Baye}, respectively. Algorithm 3 was used to construct the Bayesian credible interval.

Algorithm 3 Bayesian credible interval and HPD interval

Step 1: Specify the values of a_{i1}, a_{i2}, b_{i1}, b_{i2}, and r_i, then compute $a(r_i)$ and $b^+(r_i)$
Step 2: At the m step
(a) Generate u_i and v_i
(b) Compute $\rho_i = \frac{v_i}{u_i^{r_i}}$
(c) If the value of ρ_i is accepted, set $\beta_{i(m)} = \rho_i$ if $u_i \leqslant [p(\beta_i|x_i)]^{\frac{1}{r_i+1}}$; otherwise, repeat step (a) and step (b)
(d) Generate α_i^2 and compute $\alpha_{i(m)} = \sqrt{\alpha_i^2}$
(e) Compute θ_i, $\hat{Var}\left(\theta_{i.Baye}\right)$, and θ_{Baye}
Step 3: Repeat step 2 a total M times to obtain an array of θ_{Baye} values
Step 4: Compute L_{Baye} and U_{Baye}
Step 5: Compute L_{HPD} and U_{HPD}

2.4. Highest Posterior Density Approach

The HPD approach is a technique in Bayesian statistics used to define credible intervals (also known as Bayesian credible intervals) for estimating unknown parameters. In contrast to traditional CIs, which rely on the concept of repeated sampling, HPD intervals are directly obtained from the posterior distribution of the parameter. An HPD interval represents the smallest possible range that contains a given proportion (such as 95%) of the total posterior probability. Simply put, it identifies the most plausible range of values for the parameter, based on both the observed data and prior knowledge. In this way, the HPD interval highlights the values with the highest probability density under the posterior distribution.

The HPD interval was derived from the posterior distribution described in Equation (28). Among all intervals that include $100(1-\gamma)\%$ of the posterior probability, it has the shortest

length. Every value inside the interval has a higher posterior density than any value outside it.

Therefore, the $100(1-\gamma)\%$ two-sided HPD interval is

$$CI_{HPD} = [L_{HPD}, U_{HPD}], \tag{30}$$

where L_{HPD} and U_{HPD} are obtained using the hdi function from the HDInterval package in the R software (version 2024.12.0+467) environment. Algorithm 3 was used to construct the HPD interval.

3. Results

This study investigates various techniques for constructing CIs, including the GCI, bootstrap, Bayesian, and HPD approaches. The effectiveness of these approaches was evaluated through a Monte Carlo simulation implemented in R software (version 2024.12.0+467). The simulation aimed to assess interval estimates for the common percentile of the BS distribution. Two primary performance indicators were used in the analysis: CP and AL. CP reflects the frequency with which the interval captures the true parameter value, whereas AL serves as a measure of precision, with shorter intervals indicating greater accuracy. A CI is considered optimal if it achieves a CP of at least 0.95 and simultaneously maintains the shortest possible AL. Through this comparative analysis, the study provides valuable insights into the reliability and efficiency of each method, aiding in the selection of the most appropriate technique for estimating intervals within the BS distribution framework.

To evaluate the effectiveness of the proposed CIs, Monte Carlo simulations were carried out using R software (version 2024.12.0+467) under various conditions, including different sample sizes (n_i) and shape parameters (α_i). The scale parameter (β_i) was fixed at 1.00 across all simulation scenarios to maintain consistency. The simulation study consisted of 5000 Monte Carlo replications. In each replication, 1000 resampling iterations were conducted for each of the GCI, bootstrap, Bayesian, and HPD approaches. The performance of each method was assessed based on two key metrics: CP, defined as the proportion of times the true parameter value was captured within the CI, and AL, which reflects the precision of the interval, with shorter lengths indicating more accurate estimates. An ideal method is expected to achieve a CP equal to or exceeding the nominal confidence level of 0.95, while also minimizing the AL. The detailed computational procedure for estimating CPs and ALs for all approaches is outlined in Algorithm 4.

Algorithm 4 CPs and ALs

Step 1: Specify the values of n_i, α_i, β_i, and θ
Step 2: Generate a sample x_{ij} from the BS distributions
Step 3: Compute $\hat{\theta}_i$, $\hat{Var}(\hat{\theta}_i)$, and $\hat{\theta}$
Step 4: Apply Algorithm 1 to construct $[L_{GCI(m)}, U_{GCI(m)}]$
Step 5: Apply Algorithm 2 to construct $[L_{BCI(m)}, U_{BCI(m)}]$
Step 6: Apply Algorithm 3 to construct $[L_{Baye(m)}, U_{Baye(m)}]$ and $[L_{HPD(m)}, U_{HPD(m)}]$
Step 7: If $L_{(m)} \leqslant \theta \leqslant U_{(m)}$, set $p_{(m)} = 1$ else set $p_{(m)} = 0$
Step 8: Compute $U_{(m)} - L_{(m)}$
Step 9: Repeat step 2–step 8, a total M times
Step 10: Compute the CPs as the mean of $p_{(m)}$
Step 11: Compute the ALs as the mean of $U_{(m)} - L_{(m)}$

Tables 1 and 2 report the empirical CP and AL of four methods for constructing 95% CIs for a common percentile, evaluated through simulations under different combinations of sample sizes and parameters. For the three-sample cases, the sample size tuples (n_1, n_2, n_3)

represent the sizes of three independent samples drawn from BS distributions, with the shape parameters shown in parentheses $(\alpha_1, \alpha_2, \alpha_3)$. Based on the results presented in Tables 1 and 2, the GCI consistently has the longest ALs; however, its CPs are generally close to the nominal confidence level of 0.95 across most scenarios. In contrast, the bootstrap CI yields the shortest ALs, but its CPs fall below the nominal level of 0.95 in all cases, making it unsuitable for constructing CIs for a common percentile. The Bayesian credible interval achieves CPs near 0.95 in some scenarios, and its AL tends to be shorter than that of the GCI when both maintain similar coverage levels. However, the HPD interval consistently shows CPs below 0.95. Therefore, the GCI approach is recommended for constructing CIs for a common percentile in all cases, while the Bayesian approach becomes more suitable as sample sizes increase.

Table 1. The CPs and ALs of 95% two-sided CIs for the common percentile of BS distributions: 3 sample cases.

(n_1, n_2, n_3)	$(\alpha_1, \alpha_2, \alpha_3)$	CP (AL)			
		CI_{GCI}	CI_{BCI}	CI_{Baye}	CI_{HPD}
(10, 10, 10)	(0.25, 0.25, 0.25)	0.9382 (0.1926)	0.8874 (0.1694)	0.9270 (0.1849)	0.9188 (0.1822)
	(0.25, 0.50, 0.50)	0.9510 (0.2881)	0.9096 (0.2396)	0.9450 (0.2745)	0.9410 (0.2691)
	(0.25, 1.00, 1.00)	0.9432 (0.4762)	0.9166 (0.3841)	0.9366 (0.4483)	0.9324 (0.4367)
	(0.50, 0.50, 0.50)	0.9552 (0.3021)	0.9146 (0.2462)	0.9454 (0.2871)	0.9470 (0.2827)
	(0.50, 1.00, 1.00)	0.9474 (0.3843)	0.9152 (0.3092)	0.9402 (0.3632)	0.9342 (0.3578)
	(1.00, 1.00, 1.00)	0.9226 (0.3279)	0.8910 (0.2576)	0.9204 (0.3098)	0.9160 (0.3066)
(10, 30, 30)	(0.25, 0.25, 0.25)	0.9414 (0.1192)	0.9154 (0.1136)	0.9386 (0.1171)	0.9316 (0.1159)
	(0.25, 0.50, 0.50)	0.9420 (0.1837)	0.9170 (0.1720)	0.9368 (0.1805)	0.9310 (0.1788)
	(0.25, 1.00, 1.00)	0.9514 (0.2723)	0.9328 (0.2507)	0.9466 (0.2648)	0.9398 (0.2625)
	(0.50, 0.50, 0.50)	0.9494 (0.1787)	0.9336 (0.1622)	0.9448 (0.1739)	0.9438 (0.1720)
	(0.50, 1.00, 1.00)	0.9482 (0.2089)	0.9364 (0.1887)	0.9468 (0.2023)	0.9430 (0.2005)
	(1.00, 1.00, 1.00)	0.9416 (0.2104)	0.9168 (0.1732)	0.9374 (0.1988)	0.9436 (0.1954)
(30, 30, 30)	(0.25, 0.25, 0.25)	0.9454 (0.1034)	0.9274 (0.0998)	0.9394 (0.1019)	0.9370 (0.1009)
	(0.25, 0.50, 0.50)	0.9420 (0.1491)	0.9314 (0.1416)	0.9406 (0.1470)	0.9382 (0.1456)
	(0.25, 1.00, 1.00)	0.9462 (0.2480)	0.9352 (0.2259)	0.9442 (0.2399)	0.9408 (0.2367)
	(0.50, 0.50, 0.50)	0.9536 (0.1514)	0.9434 (0.1423)	0.9512 (0.1489)	0.9472 (0.1475)
	(0.50, 1.00, 1.00)	0.9510 (0.2001)	0.9374 (0.1807)	0.9474 (0.1934)	0.9486 (0.1914)
	(1.00, 1.00, 1.00)	0.9428 (0.1669)	0.9250 (0.1477)	0.9390 (0.1611)	0.9386 (0.1597)

Table 1. *Cont.*

(n_1, n_2, n_3)	$(\alpha_1, \alpha_2, \alpha_3)$	CP (AL)			
		CI_{GCI}	CI_{BCI}	CI_{Baye}	CI_{HPD}
(30, 50, 50)	(0.25, 0.25, 0.25)	0.9466	0.9356	0.9462	0.9430
		(0.0853)	(0.0832)	(0.0844)	(0.0836)
	(0.25, 0.50, 0.50)	0.9518	0.9400	0.9498	0.9480
		(0.1275)	(0.1231)	(0.1260)	(0.1249)
	(0.25, 1.00, 1.00)	0.9512	0.9454	0.9510	0.9446
		(0.2003)	(0.1889)	(0.1955)	(0.1937)
	(0.50, 0.50, 0.50)	0.9518	0.9454	0.9502	0.9478
		(0.1236)	(0.1183)	(0.1221)	(0.1210)
	(0.50, 1.00, 1.00)	0.9530	0.9446	0.9522	0.9472
		(0.1537)	(0.1434)	(0.1497)	(0.1484)
	(1.00, 1.00, 1.00)	0.9480	0.9340	0.9400	0.9410
		(0.1346)	(0.1221)	(0.1304)	(0.1293)
(50, 50, 50)	(0.25, 0.25, 0.25)	0.9484	0.9376	0.9456	0.9422
		(0.0791)	(0.0775)	(0.0783)	(0.0776)
	(0.25, 0.50, 0.50)	0.9546	0.94920	0.9520	0.9492
		(0.1137)	(0.1102)	(0.1127)	(0.1116)
	(0.25, 1.00, 1.00)	0.9502	0.9410	0.9448	0.9428
		(0.1867)	(0.1752)	(0.1818)	(0.1799)
	(0.50, 0.50, 0.50)	0.9490	0.9434	0.9462	0.9440
		(0.1144)	(0.1102)	(0.1132)	(0.1122)
	(0.50, 1.00, 1.00)	0.9506	0.9426	0.9490	0.9460
		(0.1499)	(0.1397)	(0.1458)	(0.1445)
	(1.00, 1.00, 1.00)	0.9436	0.9358	0.9434	0.9418
		(0.1233)	(0.1133)	(0.1199)	(0.1189)
(50, 100, 100)	(0.25, 0.25, 0.25)	0.9486	0.9404	0.9442	0.9402
		(0.0608)	(0.0599)	(0.0604)	(0.0598)
	(0.25, 0.50, 0.50)	0.9516	0.9428	0.9484	0.9450
		(0.0915)	(0.0899)	(0.0907)	(0.0900)
	(0.25, 1.00, 1.00)	0.9482	0.9412	0.9464	0.9430
		(0.1408)	(0.1359)	(0.1381)	(0.1370)
	(0.50, 0.50, 0.50)	0.9538	0.9460	0.9478	0.9468
		(0.0874)	(0.0852)	(0.0867)	(0.0859)
	(0.50, 1.00, 1.00)	0.9504	0.9440	0.9484	0.9448
		(0.1060)	(0.1015)	(0.1037)	(0.1029)
	(1.00, 1.00, 1.00)	0.9454	0.9382	0.9452	0.9422
		(0.0928)	(0.0872)	(0.0905)	(0.0898)
(100, 100, 100)	(0.25, 0.25, 0.25)	0.9482	0.9420	0.9452	0.9426
		(0.0555)	(0.0548)	(0.0551)	(0.0546)
	(0.25, 0.50, 0.50)	0.9496	0.9456	0.9458	0.9448
		(0.0797)	(0.0783)	(0.0791)	(0.0784)
	(0.25, 1.00, 1.00)	0.9438	0.9408	0.9420	0.9390
		(0.1280)	(0.1232)	(0.1254)	(0.1242)
	(0.50, 0.50, 0.50)	0.9508	0.9418	0.9454	0.9416
		(0.0795)	(0.0778)	(0.0789)	(0.0782)
	(0.50, 1.00, 1.00)	0.9470	0.9402	0.9450	0.9436
		(0.1026)	(0.0980)	(0.1004)	(0.0995)
	(1.00, 1.00, 1.00)	0.9474	0.9446	0.9484	0.9458
		(0.0837)	(0.0793)	(0.0818)	(0.0812)

Table 2. The CPs and ALs of 95% two-sided CIs for the common percentile of BS distributions: 6 sample cases.

$(n_1, n_2, n_3, n_4, n_5, n_6)$	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$	CP (AL)			
		CI_{GCI}	CI_{BCI}	CI_{Baye}	CI_{HPD}
(10, 10, 10, 10, 10, 10)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9332 (0.1375)	0.8564 (0.1246)	0.9158 (0.1320)	0.9084 (0.1305)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9478 (0.2000)	0.9004 (0.1678)	0.9408 (0.1894)	0.9354 (0.1857)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9458 (0.4280)	0.9206 (0.3097)	0.9426 (0.3867)	0.9574 (0.3631)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9412 (0.2310)	0.9094 (0.1807)	0.9376 (0.2170)	0.9420 (0.2130)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9166 (0.3499)	0.8806 (0.2558)	0.9170 (0.3223)	0.9336 (0.3120)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.8470 (0.2808)	0.7690 (0.1936)	0.8494 (0.2596)	0.8696 (0.2548)
(10, 10, 10, 30, 30, 30)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9376 (0.0925)	0.8996 (0.0885)	0.9292 (0.0905)	0.9236 (0.0896)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9466 (0.1434)	0.9046 (0.1348)	0.9408 (0.1405)	0.9346 (0.1392)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9580 (0.2346)	0.9358 (0.2123)	0.9566 (0.2270)	0.9498 (0.2245)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9494 (0.1456)	0.9332 (0.1264)	0.9444 (0.1399)	0.9462 (0.1380)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9474 (0.1801)	0.9326 (0.1571)	0.9436 (0.1731)	0.9428 (0.1714)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.8802 (0.2012)	0.8256 (0.1417)	0.8786 (0.1820)	0.9142 (0.1754)
(30, 30, 30, 30, 30, 30)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9416 (0.0732)	0.9234 (0.0713)	0.9370 (0.0722)	0.9316 (0.0716)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9436 (0.0990)	0.9314 (0.0950)	0.9416 (0.0975)	0.9390 (0.0966)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9476 (0.1759)	0.9416 (0.1569)	0.9468 (0.1687)	0.9528 (0.1653)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9504 (0.1084)	0.9408 (0.1015)	0.9470 (0.1066)	0.9438 (0.1056)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9376 (0.1562)	0.9240 (0.1373)	0.9344 (0.1501)	0.9442 (0.1479)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.9042 (0.1258)	0.8760 (0.1062)	0.9020 (0.1205)	0.9094 (0.1192)
(30, 30, 30, 50, 50, 50)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9464 (0.0629)	0.9318 (0.0616)	0.9436 (0.0622)	0.9416 (0.0617)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9468 (0.0898)	0.9298 (0.0872)	0.9422 (0.0889)	0.9376 (0.0881)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9544 (0.1509)	0.9470 (0.1408)	0.9510 (0.1465)	0.9500 (0.1449)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9404 (0.0922)	0.9332 (0.0877)	0.9382 (0.0909)	0.9382 (0.0901)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9480 (0.1235)	0.9404 (0.1133)	0.9456 (0.1198)	0.9448 (0.1187)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.9100 (0.1050)	0.8888 (0.0913)	0.9086 (0.1010)	0.9140 (0.1000)

Table 2. *Cont.*

$(n_1, n_2, n_3, n_4, n_5, n_6)$	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$	CP (AL)			
		CI_{GCI}	CI_{BCI}	CI_{Baye}	CI_{HPD}
(50, 50, 50, 50, 50, 50)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9490 (0.0559)	0.9394 (0.0550)	0.9450 (0.0554)	0.9394 (0.0549)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9456 (0.0749)	0.9390 (0.0731)	0.9418 (0.0741)	0.9408 (0.0734)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9470 (0.1262)	0.9404 (0.1175)	0.9464 (0.1224)	0.9490 (0.1208)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9536 (0.0815)	0.9432 (0.0784)	0.9496 (0.0806)	0.9476 (0.0799)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9446 (0.1137)	0.9348 (0.1045)	0.9442 (0.1102)	0.9460 (0.1090)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.9158 (0.0903)	0.8966 (0.0809)	0.9132 (0.0874)	0.9140 (0.0867)
(50, 50, 50, 100, 100, 100)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9450 (0.0454)	0.9370 (0.0449)	0.9430 (0.0451)	0.9406 (0.0447)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9506 (0.0658)	0.9470 (0.0647)	0.9476 (0.0652)	0.9466 (0.0647)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9560 (0.1066)	0.9508 (0.1024)	0.9568 (0.1044)	0.9532 (0.1035)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9464 (0.0656)	0.9420 (0.0637)	0.9456 (0.0650)	0.9432 (0.0645)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9450 (0.0851)	0.9432 (0.0808)	0.9474 (0.0832)	0.9450 (0.0825)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.9282 (0.0713)	0.9200 (0.0656)	0.9288 (0.0692)	0.9296 (0.0686)
(100, 100, 100, 100, 100, 100)	(0.25, 0.25, 0.25, 0.25, 0.25, 0.25)	0.9534 (0.0392)	0.9452 (0.0388)	0.9520 (0.0389)	0.9492 (0.0386)
	(0.25, 0.25, 0.25, 0.50, 0.50, 0.50)	0.9466 (0.0521)	0.9416 (0.0514)	0.9438 (0.0517)	0.9434 (0.0513)
	(0.25, 0.25, 0.25, 1.00, 1.00, 1.00)	0.9482 (0.0845)	0.9428 (0.0811)	0.9458 (0.0827)	0.9464 (0.0818)
	(0.50, 0.50, 0.50, 0.50, 0.50, 0.50)	0.9470 (0.0564)	0.9428 (0.0552)	0.9436 (0.0559)	0.9398 (0.0555)
	(0.50, 0.50, 0.50, 1.00, 1.00, 1.00)	0.9470 (0.0762)	0.9442 (0.0725)	0.9474 (0.0743)	0.9484 (0.0737)
	(1.00, 1.00, 1.00, 1.00, 1.00, 1.00)	0.9338 (0.0601)	0.9252 (0.0563)	0.9348 (0.0587)	0.9326 (0.0582)

4. An Empirical Application

This section focuses on illustrating the proposed inferential framework using real $PM_{2.5}$ data rather than on predictive time-series modeling. Although the observations are recorded sequentially over time, the analysis treats the data as realizations from underlying distributions to support inference on common percentiles across monitoring stations.

$PM_{2.5}$ denotes fine particles with a diameter of 2.5 micrometers or smaller. Due to their tiny size, they can reach deep into the lungs and potentially enter the bloodstream, posing significant health hazards. Percentiles are statistical tools used to illustrate the variation in $PM_{2.5}$ levels over time. For example, the 50th percentile, or median, indicates the value below which half of the $PM_{2.5}$ readings fall. Likewise, the 95th percentile shows that 95% of the values are below a certain level, while the remaining 5% exceed it. Compared to simple averages, percentiles provide more detailed insight into pollution trends. They are

particularly valuable in environmental and public health research, especially for detecting periods of high pollution. Using percentiles enables researchers and policymakers to monitor changes in air quality, identify extreme pollution events, and compare pollution levels across different regions and time periods. Therefore, percentile-based analysis of $PM_{2.5}$ offers a more accurate and comprehensive understanding of air quality and its impact on human health and the environment.

Table 3 presents $PM_{2.5}$ levels recorded at six monitoring stations in Chiang Mai Province between 1 January and 30 April 2025, as reported by the Pollution Control Department (http://air4thai.pcd.go.th/webV3/#/History, accessed on 1 November 2025). The stations are located in the following areas: Station 1—Chang Phueak Subdistrict, Mueang Chiang Mai District; Station 2—Si Phum Subdistrict, Mueang Chiang Mai District; Station 3—Suthep Subdistrict, Mueang Chiang Mai District; Station 4—Chang Khoeng Subdistrict, Mae Chaem District; Station 5—Mueang Na Subdistrict, Chiang Dao District; and Station 6—Hang Dong Subdistrict, Hot District. The $PM_{2.5}$ measurements considered in this study are obtained from secondary data sources. Despite the presence of an N/A category denoting unavailable data, a detailed review of the dataset confirmed the absence of such entries throughout the study period, eliminating the need for any imputation or data correction procedures.

Figures 1–6 show histograms illustrating the distribution of daily $PM_{2.5}$ level data collected from six monitoring stations in Chiang Mai Province. These figures reveal that the data exhibit positively skewed distributions across all stations. The time-series plots and histograms are presented to describe the empirical characteristics of the $PM_{2.5}$ data and to motivate the distributional assumptions, not to model temporal dependence or forecast future concentrations. Since the proposed analysis does not model temporal dependence or perform forecasting, stationarity assumptions and related tests are not required. Table 4 presents the estimated Akaike Information Criterion (AIC) values for various probability models fitted to the daily $PM_{2.5}$ level data, including the normal, log-normal, Weibull, gamma, exponential, logistic, Cauchy, and BS distributions. The results indicate that, for all six stations, the BS distribution yields the lowest AIC values. This suggests that the BS distribution is the most suitable model for representing the daily $PM_{2.5}$ level data in Chiang Mai Province. Although the BS distribution consistently produces the lowest AIC values across all six stations and the histograms exhibit comparable right-skewed patterns, this evidence is suggestive rather than definitive. Nonetheless, it lends support to the plausibility of a shared-quantile assumption across sites. Distributional assumptions are evaluated using histograms and empirical cumulative distribution functions rather than normality tests, as the analysis is based on the BS distribution and does not assume normality. Since the proposed analysis does not involve time-series modeling or the specification of autoregressive or seasonal structures, autocorrelation diagnostics such as ACF and PACF are not considered.

Table 5 presents the sample statistics of daily $PM_{2.5}$ level data collected from six monitoring stations in Chiang Mai Province. The estimated common percentile across these stations is $\hat{\theta} = 18.1350$. Accordingly, the fitted BS model is used as a working distribution to estimate shared percentiles and to construct CIs, thereby demonstrating the practical applicability of the proposed inferential methods. Since the objective of this study is statistical inference rather than prediction, the empirical analysis focuses on parameter estimation and CI construction. Table 6 summarizes the 95% CIs for the common percentile obtained using four inferential approaches. Table 6 shows the 95% two-sided CIs for this common percentile, based on the BS distribution and calculated using four approaches: GCI, bootstrap, Bayesian, and HPD approaches. The best-performing approach is highlighted in bold. Among these, the HPD approach produced the shortest interval length for the daily $PM_{2.5}$ data, while the bootstrap approach resulted in the longest. However, in simulation studies using 5000 random

samples, the bootstrap intervals were shorter than those from the GCI, Bayesian, and HPD approaches, but their CPs fell below the nominal level of 0.95. In contrast, the example-based interval lengths were derived from a single sample. Based on both CP and AL, the GCI approach is recommended for constructing CIs for the common percentile.

Table 3. Daily $PM_{2.5}$ levels data of six stations in Chiang Mai Province.

Stations	**$PM_{2.5}$ Levels**									
Station 1	25.1	21.6	20.4	22.4	25.3	25.3	23.2	29.6	31.5	25.8
	26.4	22.4	21.0	23.3	35.5	29.8	26.7	23.4	22.0	18.5
	16.5	25.1	27.4	29.3	35.4	34.2	33.4	32.8	32.9	33.5
	28.4	19.6	16.8	16.8	33.0	39.1	32.2	28.2	31.4	39.0
	39.0	45.2	40.1	39.2	35.3	34.0	32.7	37.4	34.9	26.1
	30.6	26.4	26.0	43.9	35.4	29.8	33.1	54.1	56.3	48.3
	39.9	47.6	42.7	40.3	39.1	63.5	86.0	101.9	70.3	56.9
	66.8	64.1	51.6	56.4	118.2	143.6	106.0	88.7	112.6	97.2
	30.1	33.6	39.3	41.4	59.1	53.5	75.9	67.2	77.3	75.3
	97.1	87.9	100.7	108.2	123.9	106.5	134.1	158.2	138.0	122.7
	102.2	46.6	41.2	43.8	52.2	61.9	49.3	49.0	76.4	97.8
	89.4	72.4	59.1	73.8	69.0	70.9	94.2	78.7	76.0	74.9
	83.4									
Station 2	28.7	19.8	18.6	18.4	21.8	20.5	19.5	26.6	27.3	23.3
	20.8	22.0	24.6	20.0	32.4	24.7	21.9	17.9	19.6	15.6
	16.9	23.2	28.9	30.5	37.9	33.7	29.5	30.3	34.7	27.3
	23.8	14.2	16.8	16.5	32.4	43.9	33.8	31.4	30.3	43.9
	38.3	32.0	30.5	39.6	30.0	35.6	31.9	38.4	34.2	23.5
	26.4	21.9	26.4	38.6	37.3	35.1	34.6	51.3	54.0	45.2
	41.4	53.8	45.5	41.8	40.0	62.8	103.9	115.1	90.5	57.9
	69.9	74.9	52.2	56.6	122.6	133.4	106.3	91.1	125.0	85.0
	37.6	36.0	40.6	43.8	65.8	57.5	93.6	70.1	71.9	85.4
	95.8	93.8	99.8	111.3	134.1	111.0	131.7	157.3	128.2	129.8
	99.3	41.8	36.5	39.7	44.8	60.3	47.6	47.4	75.3	86.2
	95.8	77.9	56.5	70.3	61.8	61.8	77.2	73.8	71.5	69.3
	79.4									
Station 3	12.2	11.3	11.7	13.9	15.2	15.0	22.2	21.4	17.0	15.8
	13.2	12.5	20.2	24.6	19.4	16.9	16.4	12.9	9.7	8.8
	15.6	19.2	26.7	28.4	25.8	16.5	22.7	19.0	20.0	15.1
	8.4	7.2	12.2	31.3	27.5	21.3	20.6	26.9	30.1	35.1
	37.2	33.4	32.8	28.2	30.0	29.7	38.8	27.6	16.3	18.7
	17.1	21.1	34.8	22.6	17.8	28.8	48.8	56.4	39.4	33.2
	38.9	31.0	32.0	34.2	66.3	84.6	88.6	48.1	48.5	43.4
	34.7	34.8	58.3	136.6	125.0	89.0	72.0	93.6	18.5	26.7
	30.1	37.5	47.9	49.8	67.8	60.3	78.9	56.6	92.5	88.6
	117.8	106.0	113.8	110.9	123.7	155.3	135.7	125.5	85.1	29.3
	27.6	35.9	45.0	51.2	40.2	47.7	60.8	78.1	60.8	42.5
	58.0	61.5	59.5	61.8	58.6	57.8	67.2	63.3		
Station 4	16.2	11.1	11.5	12.8	16.2	16.3	15.1	16.0	15.3	15.9
	17.5	19.4	11.4	9.4	20.1	20.3	14.9	14.1	13.0	12.8
	11.4	21.4	22.3	20.4	18.3	26.8	21.7	16.1	22.3	22.4
	15.9	9.9	7.8	6.4	25.0	23.2	21.9	18.5	23.4	34.0
	26.2	25.2	33.8	25.5	17.9	19.3	15.2	24.4	26.0	14.6
	20.1	24.2	15.0	23.0	18.0	19.5	39.2	58.5	54.2	45.0
	48.6	59.1	38.6	30.6	28.2	42.5	50.3	54.3	35.0	44.5
	39.9	28.2	37.7	45.9	86.4	109.7	74.2	60.7	50.0	55.3
	25.7	31.3	36.5	48.8	62.1	44.4	56.2	53.5	68.9	69.8
	82.7	85.1	103.1	87.8	109.7	125.6	98.8	76.0	83.3	37.9
	29.9	30.7	40.2	39.3	35.7	57.2	78.7	89.4	93.9	69.1
	48.1	54.8	56.5	46.6	49.1	47.0	59.0	55.9	52.0	

Table 3. *Cont.*

Stations	$PM_{2.5}$ Levels									
Station 5	22.2	12.3	13.7	11.9	13.5	15.6	14.7	17.5	16.3	13.4
	17.2	14.7	12.5	14.6	23.7	20.6	15.2	14.3	18.0	12.6
	12.3	13.7	21.1	21.4	31.2	28.4	28.5	20.2	24.3	22.1
	17.8	11.1	9.1	11.6	31.5	36.4	21.5	24.1	28.2	32.3
	35.3	36.9	24.7	29.9	25.9	28.5	31.4	29.7	21.7	23.0
	28.6	24.9	27.0	31.1	36.3	39.6	37.7	52.6	49.6	37.8
	31.5	43.7	40.0	35.5	40.2	88.9	127.5	157.0	101.0	56.4
	65.4	76.2	47.7	58.4	137.1	119.4	127.9	121.5	161.4	126.6
	32.5	29.8	28.6	41.8	51.2	73.1	117.2	111.6	121.0	119.7
	144.6	155.5	216.8	198.5	200.5	175.4	224.3	209.9	168.5	151.6
	155.9	55.2	50.3	55.2	55.8	58.5	49.0	46.2	74.1	63.7
	78.1	66.6	59.4	109.2	82.4	57.0	55.4	51.9	68.7	102.7
	108.6									
Station 6	25.1	16.6	17.7	16.1	18.2	21.6	23.3	25.1	21.0	16.7
	16.9	15.3	16.0	19.1	22.9	22.8	19.1	16.2	16.1	12.7
	18.6	22.3	27.7	21.2	32.1	33.6	27.6	27.0	26.1	19.4
	11.7	11.7	13.5	23.2	35.8	27.0	26.7	29.6	38.7	33.3
	42.9	37.0	35.7	30.2	38.2	35.7	44.9	41.3	30.5	37.4
	37.6	35.7	43.3	55.1	75.1	82.4	70.7	69.0	45.7	36.7
	44.0	43.0	34.7	25.1	64.8	71.5	82.5	56.0	48.3	59.0
	48.4	42.6	37.6	98.7	147.7	110.0	88.9	68.7	63.9	24.1
	34.7	43.9	46.3	74.5	119.6	111.5	100.5	162.4	90.1	98.8
	85.8	90.5	84.1	93.7	84.1	132.9	150.1	105.5	75.3	94.6
	41.2	30.1	34.5	40.6	42.8	38.0	46.1	71.3	77.4	61.3
	52.8	53.1	54.6	51.7	49.9	43.6	47.5	56.5	53.3	55.1

Source: Pollution Control Department (http://air4thai.pcd.go.th/webV3/#/History, accessed on 1 November 2025).

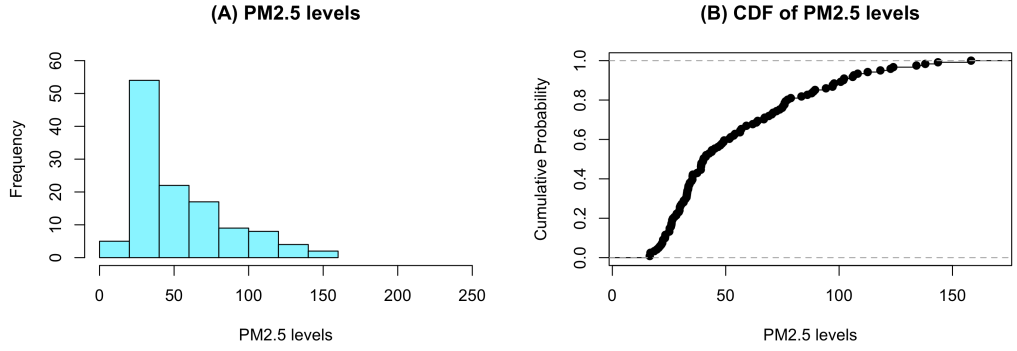


Figure 1. Histogram (**A**) and cumulative distribution function (CDF) (**B**) of daily $PM_{2.5}$ levels data for Station 1.

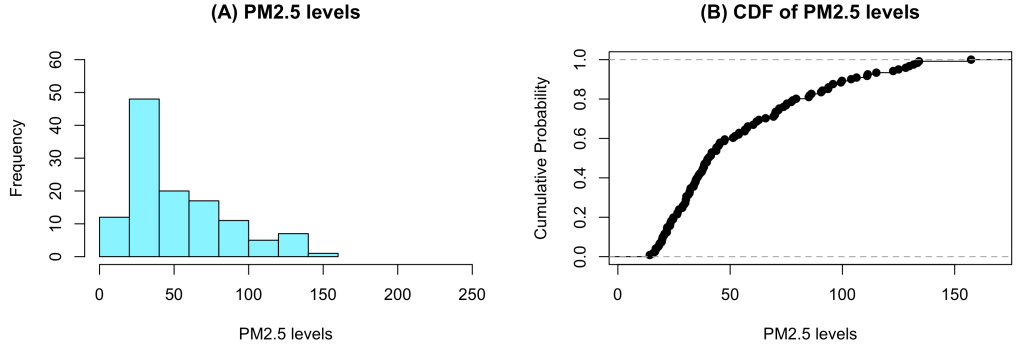


Figure 2. Histogram (**A**) and CDF (**B**) of daily $PM_{2.5}$ levels data for Station 2.

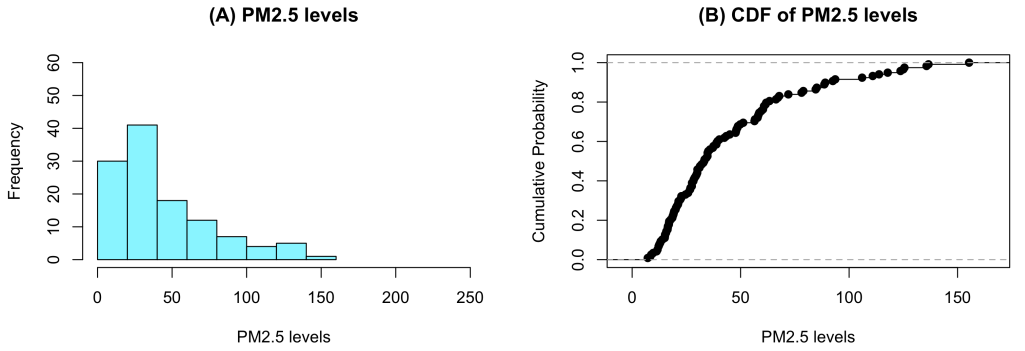


Figure 3. Histogram (**A**) and CDF (**B**) of daily $PM_{2.5}$ levels data for Station 3.

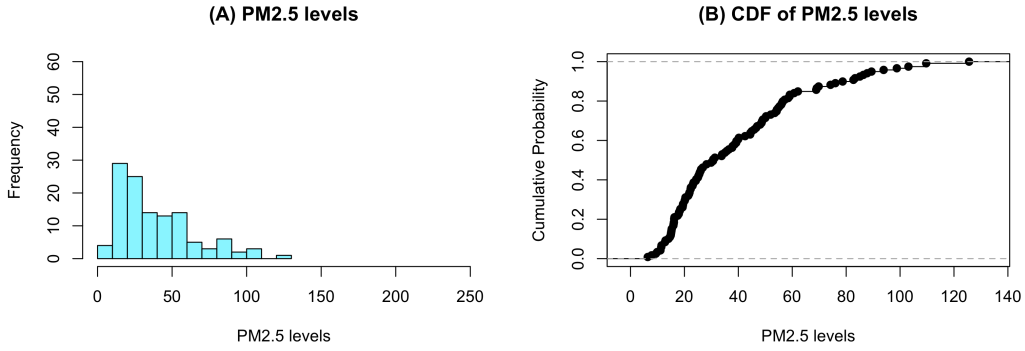


Figure 4. Histogram (**A**) and CDF (**B**) of daily $PM_{2.5}$ levels data for Station 4.

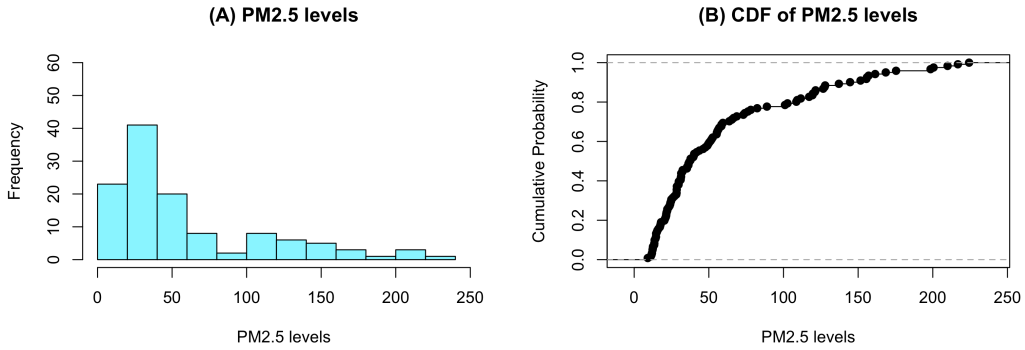


Figure 5. Histogram (**A**) and CDF (**B**) of daily $PM_{2.5}$ levels data for Station 5.

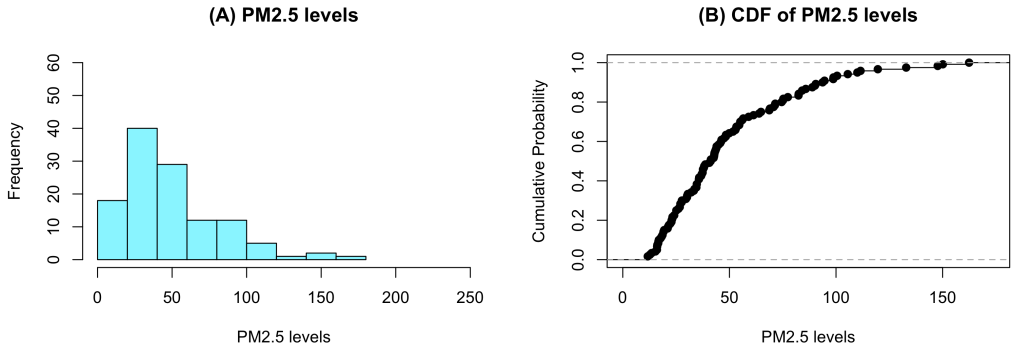


Figure 6. Histogram (**A**) and CDF (**B**) of daily $PM_{2.5}$ levels data for Station 6.

Table 4. The estimated AIC values for the probability models using daily $PM_{2.5}$ levels data from six stations in Chiang Mai Province.

Distributions	Stations					
	Station 1	**Station 2**	**Station 3**	**Station 4**	**Station 5**	**Station 6**
Normal distribution	1185.18	1194.21	1161.23	1116.59	1306.07	1173.40
Log-normal distribution	1131.93	1142.06	1094.12	1066.78	1208.59	1119.85
Weibull distribution	1152.28	1158.75	1109.09	1077.35	1227.65	1134.86
Gamma distribution	1141.22	1149.88	1102.16	1071.23	1222.82	1126.03
Exponential distribution	1208.73	1206.87	1133.60	1113.57	1234.59	1178.89
Logistic distribution	1183.17	1193.52	1153.43	1114.26	1298.07	1166.23
Cauchy distribution	1201.46	1213.32	1162.79	1143.70	1286.95	1183.19
Birnbaum–Saunders distribution	**1130.67**	**1140.47**	**1093.05**	**1065.69**	**1203.67**	**1119.53**

Note: Bold font means the distribution with the lowest AIC value.

Table 5. Sample statistics of daily $PM_{2.5}$ levels data for six stations in Chiang Mai Province.

Statistics	Stations					
	Station 1	**Station 2**	**Station 3**	**Station 4**	**Station 5**	**Station 6**
n_i	121	121	118	119	121	120
α_i	0.6126	0.6575	0.7747	0.6911	0.9833	0.6593
β_i	56.7104	55.5136	42.7081	30.7391	62.8481	47.8938
$\hat{\theta}_i$	26.3577	24.4623	16.4183	13.0209	19.1411	21.0598
$\hat{Var}(\hat{\theta}_i)$	3.4907	3.4052	2.0795	1.0693	4.0161	2.5570

Table 6. The 95% two-sided CIs for the common percentile of BS distributions, based on daily $PM_{2.5}$ level data from six stations in Chiang Mai Province.

Approaches	Confidence Intervals		
	Lower	**Upper**	**Length**
GCI	15.2117	17.5398	2.3281
Bootstrap	15.3316	17.6617	2.3301
Bayesian	15.3155	17.6143	2.2988
HPD	**15.2638**	**17.5398**	**2.2760**

Note: Bold font indicates the approach with the shortest length.

5. Discussion

In environmental research and air quality monitoring, percentiles are key to interpreting $PM_{2.5}$ concentration patterns. Unlike mean values, percentiles such as the 50th, 75th, and 90th better illustrate how pollutant levels vary over time and more accurately represent typical population exposure. Incorporating CIs into percentile estimates further strengthens their value by reflecting the uncertainty inherent in sample data, thus enabling more informed and reliable decisions. This is particularly important in public health and environmental regulation, where policies often hinge on exposure thresholds. Using percentiles together with CIs allows policymakers and environmental authorities to more effectively track pollution trends, pinpoint high-risk regions, and implement targeted measures to reduce $PM_{2.5}$ levels.

In many applications, percentiles may be more informative than variance. To address this, Thangjai et al. [15] developed the GCI approach for estimating differences in percentiles of BS distributions, using $PM_{2.5}$ data from Thailand as an example. Building on these efforts, the present study aims to estimate the common percentiles of the BS distribution.

To that end, four approaches—GCI, bootstrap, Bayesian, and HPD—were used to construct CIs for the common percentiles. The GCI approach demonstrated especially strong performance among these. Simulation data were used by all four methods. Specifically, the GCI approach employs GPQs, the bootstrap technique uses sampling distributions, and both the Bayesian and HPD approaches rely on prior distributions.

Despite its contributions, this study has several methodological limitations that warrant consideration. First, the proposed framework assumes that the underlying distribution follows a BS model; deviations from this assumption may affect the accuracy and reliability of the resulting CIs. Second, the simulation study explores a restricted range of parameter configurations and sample sizes, which may limit the robustness of the conclusions under alternative modeling scenarios. Third, the Bayesian and HPD approaches are inherently sensitive to prior specification, yet only commonly adopted priors were examined, potentially influencing posterior inference. Finally, the comparative performance of the proposed methods is evaluated under controlled simulation settings, and their behavior under more complex dependence structures or model misspecification remains an open issue for future research.

The results show that the GCI approach is the most dependable for constructing CIs for the common percentile of the BS distribution. This conclusion is supported by previous research, including the works of Thangjai et al. [15] and Ye et al. [22].

Although the GCI method is recommended due to its favorable coverage accuracy and interval precision, it is also computationally efficient. In contrast to Bayesian approaches, which typically require iterative sampling procedures such as MCMC and careful prior specification, the GCI method relies on simulation from known distributions and avoids convergence diagnostics. As a result, GCI generally entails lower computational cost and simpler implementation, particularly for large datasets or repeated analyses, while still providing reliable inferential performance. The inferior coverage performance of the bootstrap approach can be attributed to the presence of skewed and heavy-tailed distributions, under which standard bootstrap CIs may not adequately capture tail uncertainty. Moreover, when estimating a common percentile across multiple populations, the bootstrap approach may fail to properly account for the compounded uncertainty arising from parameter estimation and population heterogeneity. These factors can lead to systematic undercoverage, particularly in small to moderate sample sizes.

Several promising avenues can extend the present study and situate it within a broader interdisciplinary context. First, applications of causal inference in health outcomes, particularly using instrumental-variable frameworks, can provide robust insights into treatment effects and policy interventions (Nan and Jay [23]). Second, advanced time-series modeling approaches for $PM_{2.5}$ forecasting—integrating both theoretical framework development and empirical validation—offer opportunities to improve predictive accuracy and support public health decision-making (Wu et al. [24]). Third, exploring the joint effects of multiple air pollutants, such as $PM_{2.5}$ and O_3, on respiratory and cardiovascular outcomes can elucidate complex exposure-response relationships (Li et al. [25]). Beyond environmental health, statistical modeling approaches for optimizing energy-storage capacity in renewable energy systems present a valuable avenue for sustainability research (Zhang et al. [26]). Finally, emerging reliability estimation techniques based on nonlinear Tweedie exponential dispersion processes combined with evidential reasoning provide a rigorous framework for analyzing complex industrial and engineering systems (Lu et al. [27]). Pursuing these directions would not only extend the current work but also strengthen its relevance across environmental science, public health, and engineering disciplines.

6. Conclusions

This study developed confidence intervals for the common percentile of BS distributions using the GCI, bootstrap, Bayesian, and HPD approaches. Simulation results indicate that although the bootstrap, Bayesian, and HPD methods often produce shorter intervals, they tend to suffer from undercoverage, reflecting sensitivity to distributional asymmetry and parameter uncertainty. In contrast, the GCI approach consistently achieves coverage probabilities at or above the nominal level and yields competitive interval lengths, demonstrating robustness to skewness and asymmetry across multiple populations. These findings highlight the GCI method as a stable and symmetry-preserving inferential framework for common percentile estimation, making it particularly suitable for applications involving heterogeneous and asymmetric data structures modeled by BS distributions.

Author Contributions: Conceptualization, S.-A.N. and W.T.; methodology, S.-A.N. and W.T.; software, W.T. and R.P.; validation, S.-A.N. and S.N.; formal analysis, S.-A.N. and W.T.; investigation, S.N. and R.P.; resources, W.T.; data curation, W.T.; writing—original draft preparation, W.T.; writing—review and editing, S.-A.N. and W.T.; visualization, W.T.; supervision, S.-A.N.; project administration, S.-A.N. and S.N.; funding acquisition, S.-A.N. and S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research budget was allocated by National Science, Research, and Innovation Fund (NSRF) and King Mongkut's University of Technology North Bangkok: KMUTNB-FF-68-B-43.

Data Availability Statement: Daily $PM_{2.5}$ level data were obtained from the Pollution Control Department and are available on the website at http://air4thai.pcd.go.th/webV3/#/History, (accessed on 1 November 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Suppose that $X = (X_1, X_2, \ldots, X_n)$ is a random variable drawn from BS distribution with parameters α and β. Let θ be the percentile of X and let $\hat{\theta}$ be an estimator of θ. The variance of $\hat{\theta}$ is

$$\begin{aligned} Var(\hat{\theta}) &= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2 \\ &* \left(\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\alpha^2}{2(2+\alpha^2)^2}\right)\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\right). \end{aligned}$$

Proof. The percentile of X is defined as

$$f(\alpha, \beta) = \theta = \frac{\beta}{4}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2.$$

The estimator of $f(\alpha, \beta)$ is obtained through a Taylor series expansion, expressed as

$$f(\hat{\alpha}, \hat{\beta}) \approx f(\alpha, \beta) + \frac{\partial f(\alpha, \beta)}{\partial \alpha}(\hat{\alpha} - \alpha) + \frac{\partial f(\alpha, \beta)}{\partial \beta}(\hat{\beta} - \beta).$$

Here, $\frac{\partial f(\alpha,\beta)}{\partial \alpha}$ represents the derivative of $f(\alpha, \beta)$ evaluated at the point α. Consequently, $\frac{\partial f(\alpha,\beta)}{\partial \alpha}$ can be determined as follows

$$\begin{aligned}\frac{\partial f(\alpha,\beta)}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\frac{\beta}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2\right)\\ &= \frac{\beta}{4}\left(2\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\frac{1}{2}\left(\alpha^2 z_p^2+4\right)^{-\frac{1}{2}}\left(2z_p^2\alpha\right)\right)\right)\\ &= \frac{\beta}{2}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\alpha z_p^2\left(\alpha^2 z_p^2+4\right)^{-\frac{1}{2}}\right)\\ &= \frac{\beta}{2}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2+4}}\right).\end{aligned}$$

Likewise, $\frac{\partial f(\alpha,\beta)}{\partial \beta}$ represents the derivative of $f(\alpha,\beta)$ evaluated at the point β. Hence, $\frac{\partial f(\alpha,\beta)}{\partial \beta}$ can be obtained as follows

$$\begin{aligned}\frac{\partial f(\alpha,\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta}\left(\frac{\beta}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2\right)\\ &= \frac{1}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2.\end{aligned}$$

Therefore, the estimator of $f(\alpha,\beta)$ is

$$\begin{aligned}f(\hat{\alpha},\hat{\beta}) &\approx f(\alpha,\beta)+\frac{\partial f(\alpha,\beta)}{\partial \alpha}(\hat{\alpha}-\alpha)+\frac{\partial f(\alpha,\beta)}{\partial \beta}(\hat{\beta}-\beta)\\ &= \frac{\beta}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2+\frac{\beta}{2}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2+4}}\right)(\hat{\alpha}-\alpha)\\ &+ \frac{1}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2(\hat{\beta}-\beta).\end{aligned}$$

The variance of $\hat{\theta}$ can be estimated using the delta method, given by

$$\begin{aligned}Var(\hat{\theta}) &= Var(f(\hat{\alpha},\hat{\beta}))\\ &\approx Var\left(f(\alpha,\beta)+\frac{\partial f(\alpha,\beta)}{\partial \alpha}(\hat{\alpha}-\alpha)+\frac{\partial f(\alpha,\beta)}{\partial \beta}(\hat{\beta}-\beta)\right)\\ &= Var\left(\frac{\beta}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2+\frac{\beta}{2}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2+4}}\right)(\hat{\alpha}-\alpha)\right.\\ &+ \left.\frac{1}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2(\hat{\beta}-\beta)\right)\\ &= Var\left(\frac{\beta}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2\right)\\ &+ Var\left(\frac{\beta}{2}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)\left(z_p+\frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2+4}}\right)(\hat{\alpha}-\alpha)\right)\\ &+ Var\left(\frac{1}{4}\left(\alpha z_p+\sqrt{\alpha^2 z_p^2+4}\right)^2(\hat{\beta}-\beta)\right)\end{aligned}$$

$$
\begin{aligned}
Var(\hat{\theta}) &= 0 + \left(\frac{\beta}{2}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)\right)^2 Var(\hat{\alpha}) + 0 \\
&+ \left(\frac{1}{4}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\right)^2 Var(\hat{\beta}) + 0 \\
&= \frac{\beta^2}{4}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 Var(\hat{\alpha}) \\
&+ \frac{1}{16}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4 Var(\hat{\beta})
\end{aligned}
$$

Since $Var(\hat{\alpha}) = \frac{\alpha^2}{2n}$ and $Var(\hat{\beta}) = \frac{\alpha^2\beta^2}{n}\left(\frac{1+\frac{3}{4}\alpha^2}{\left(1+\frac{1}{2}\alpha^2\right)^2}\right)$. Therefore, the variance of $\hat{\theta}$ is

$$
\begin{aligned}
Var(\hat{\theta}) &= \frac{\beta^2}{4}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2\left(\frac{\alpha^2}{2n}\right) \\
&+ \frac{1}{16}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4\left(\frac{\alpha^2\beta^2}{n}\right)\left(\frac{1+\frac{3}{4}\alpha^2}{\left(1+\frac{1}{2}\alpha^2\right)^2}\right) \\
&= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 \\
&+ \frac{\alpha^2\beta^2}{16n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4\left(\frac{1+\frac{3}{4}\alpha^2}{\left(1+\frac{1}{2}\alpha^2\right)^2}\right) \\
&= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 \\
&+ \frac{\alpha^2\beta^2}{16n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4\left(\frac{\frac{4}{4}+\frac{3}{4}\alpha^2}{\left(\frac{2}{2}+\frac{1}{2}\alpha^2\right)^2}\right) \\
&= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 \\
&+ \frac{\alpha^2\beta^2}{16n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4\left(\frac{\frac{1}{4}(4+3\alpha^2)}{\left(\frac{1}{2}\right)^2(2+\alpha^2)^2}\right) \\
&= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 \\
&+ \frac{\alpha^2\beta^2}{16n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4\left(\frac{\frac{1}{4}(4+3\alpha^2)}{\frac{1}{4}(2+\alpha^2)^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{\theta}) &= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2 \left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 \\
&+ \frac{\alpha^2\beta^2}{16n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^4 \left(\frac{4 + 3\alpha^2}{(2+\alpha^2)^2}\right) \\
&= \frac{\alpha^2\beta^2}{8n}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2 \\
&* \left(\left(z_p + \frac{\alpha z_p^2}{\sqrt{\alpha^2 z_p^2 + 4}}\right)^2 + \left(\frac{4 + 3\alpha^2}{2(2+\alpha^2)^2}\right)\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2\right).
\end{aligned}
$$

□

References

1. Marshall, A.W.; Walsh, J.E. Some test for comparing percentage points of two arbitrary continuous populations. In *Proceedings of the International Congress of Mathematicians*; American Mathematical Society: Providence, RI, USA,1950; Volume I, pp. 1–13.
2. Harrell, F.E.; Davis, C.E. A new distribution-free quantile estimator. *Biometrika* **1982**, *69*, 635–640. [CrossRef]
3. Kaigh, W.D.; Lachenbruch, P.A. A generalized quantile estimator. *Commun. Stat. Theory Methods* **1982**, *11*, 2217–2238.
4. Albers, W.; Löhnberg, P. An approximate confidence interval for the difference between quantiles in a biomedical problem. *Stat. Neerl.* **1984**, *38*, 20–22.
5. Cox, T.F.; Jaber, K. Testing the equality of two normal percentiles. *Commun. Stat. Simul. Comput.* **1985**, *14*, 345–356.
6. Chang, D.S.; Tang, L.C. Percentile bounds and tolerance limits for the Birnbaum–Saunders distribution. *Commun. Stat. Theory Methods* **1994**, *23*, 2853–2863. [CrossRef]
7. Padgett, W.J.; Tomlinson, M.A. Lower confidence bounds for percentiles of Weibull and Birnbaum–Saunders distributions. *J. Stat. Comput. Sim.* **2003**, *73*, 429–443.
8. Guo, H.; Krishnamoorthy, K. Comparison between two quantiles: The normal and exponential cases. *Commun. Stat. Simul. Comput.* **2005**, *34*, 243–252. [CrossRef]
9. Huang, L.-F.; Johnson, R.A. Confidence regions for the ratio of percentiles. *Statist. Probab. Lett.* **2006**, *76*, 384–392.
10. Navruz, G.; Özdemir, A.F. Quantile estimation and comparing two independent groups with an approach based on percentile bootstrap. *Commun. Stat. Simul. Comput.* **2018**, *47*, 2119–2138.
11. Hasan, M.S.; Krishnamoorthy, K. Confidence intervals for the mean and a percentile based on zero-inflated lognormal data. *J. Stat. Comput. Sim.* **2018**, *88*, 1499–1514.
12. Abdollahnezhad, K.; Jafari, A.A. Testing the equality of quantiles for several normal populations. *Commun. Stat. Simul. Comput.* **2018**, *47*, 1890–1898.
13. Guo, X.; Wu, H.; Li, G.; Li, Q. Inference for the common mean of several Birnbaum–Saunders populations. *J. Appl. Stat.* **2017**, *44*, 941–954.
14. Shakil, M.; Munir, M.; Kausar, N.; Ahsanullah, M.; Khadim, A.; Sirajo, M.; Singh, J.N.; Kibria, M.G. Some inferences on three parameters Birnbaum–Saunders distribution: Statistical properties, characterizations and applications. *Comput. J. Math. Stat. Sci.* **2023**, *2*, 197–222. [CrossRef]
15. Thangjai, W.; Niwitpong, S.-A.; Niwitpong, S. Generalized confidence interval for the difference between percentiles of Birnbaum–Saunders distributions and its application to PM2.5 in Thailand. *Comput. Math. Methods* **2024**, *2024*, 2599243. [CrossRef]
16. Leiva, V.; Barros, M.; Paula, G.A.; Sanhueza, A. Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. *Environmetrics* **2008**, *19*, 235–249. [CrossRef]
17. Wang, B.X. Generalized interval estimation for the Birnbaum–Saunders distribution. *Comput. Stat. Data Anal.* **2012**, *56*, 4320–4326. [CrossRef]
18. Niu, C.; Guo, X.; Xu, W.; Zhu, L. Comparison of several Birnbaum–Saunders distributions. *J. Stat. Comput. Simulation* **2014**, *84*, 2721–2733.
19. Sun, Z.L. The confidence intervals for the scale parameter of the Birnbaum–Saunders fatigue life distribution. *Acta Armament* **2009**, *30*, 1558–1561.
20. Wang, M.; Sun, X.; Park, C. Bayesian analysis of Birnbaum–Saunders distribution via the generalized ratio-of-uniforms method. *Comput. Stat.* **2016**, *31*, 207–225.

21. Wakefield, J.C.; Gelfand, A.E.; Smith, A.F.M. Efficient generation of random variates via the ratio-of-uniforms method. *Stat. Comp.* **1991**, *1*, 129–133. [CrossRef]
22. Ye, R.D.; Ma, T.F.; Wang, S.G. Inferences on the common mean of several inverse Gaussian populations. *Comput. Stat. Data Anal.* **2010**, *54*, 906–915. [CrossRef]
23. Nan, C.; Jay, P. The causal effect of delivery volume on severe maternal morbidity: An instrumental variable analysis in Sichuan, China. *BMJ Glob. Health* **2022**, *7*, e8428.
24. Wu, C.; Wang, R.; Lu, S.; Tian, J.; Yin, L.; Wang, L.; Zheng, W. Time-series data-driven PM2.5 forecasting: From theoretical framework to empirical analysis. *Atmosphere* **2025**, *16*, 292. [CrossRef]
25. Li, J.; Liang, L.; Lyu, B.; Cai, Y.S.; Zuo, Y.; Su, J.; Tong, Z. Double trouble: The interaction PM2.5 and O3 on respiratory hospital admissions. *Environ. Pollut.* **2023**, *338*, 122665. [CrossRef] [PubMed]
26. Zhang, Y.; Zhang, Y.; Zheng, B.; Cui, H.; Qi, H. Statistical analysis for estimating the optimized battery capacity for roof-top PV energy system. *Renew. Energy* **2025**, *242*, 122491. [CrossRef]
27. Lu, Y.; Wang, S.P.; Chen, R.; Zhang, C.; Zhang, Y.; Gao, J.; Du, S. Reliability estimation method based on nonlinear Tweedie exponential dispersion process and evidential reasoning rule. *Comput. Ind. Eng.* **2025**, *206*, 111205. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.