



Estimation methods for the ratio of medians of three-parameter lognormal distributions containing zero values and their application to wind speed data from northern Thailand

Patcharee Maneerat¹, Pisit Nakjai² and Sa-Aat Niwitpong³

¹Department of Applied Mathematics, Uttaradit Rajabhat University, Uttaradit, Thailand

²Department of Computer Sciences, Uttaradit Rajabhat University, Uttaradit, Thailand

³Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

ABSTRACT

Wind speed has an important impact on the formation and dispersion of fine particulate matter (PM), which can cause several health problems. During the transition from the winter to the summer season in northern Thailand, the wind speed has been low for longer than usual, which has resulted in fine PM accumulating in the air. Motivated by this, we have identified a need to investigate wind speed due to its effect on PM formation and dispersion and to raise awareness among the general public. The hourly windspeed can be approximated by using confidence intervals for the ratio of the medians of three-parameter lognormal distributions containing zero values. Thus, we constructed them by using fiducial, normal approximation, and Bayesian methods. By way of comparison, the performance measures for all of the proposed methods (the coverage percentage, lower and upper error probabilities (LEP and UEP, respectively), and expected length) were assessed *via* Monte Carlo simulation. The results of Monte Carlo simulation studies show that the Bayesian method provided coverage percentages close to the nominal confidence level and shorter intervals than the other methods. Importantly, it maintained a good balance between LEP and UEP even for large variation and percentage of zero-valued observations. To illustrate the efficacy of our proposed methods, we applied them to hourly wind speed data from northern Thailand.

Subjects Statistics, Computational Science, Natural Resource Management, Environmental Contamination and Remediation, Environmental Impacts

Keywords Wind speed, Particulate matter, Confidence intervals, Ratio of medians, Fiducial, Normal approximation, Bayesian method

INTRODUCTION

Oxygen in the air is necessary for the survival of humans and other animals. How fast the air moves past a certain point is known as wind speed (measured in km/h), which is an important phenomenon in meteorology for monitoring and predicting weather patterns in a given area. It is reported daily along with the temperature, precipitation, and humidity for all provinces on the Thai Meteorological Department website. Importantly, low wind speeds

Submitted 16 June 2022
Accepted 15 September 2022
Published 11 October 2022

Corresponding author
Sa-Aat Niwitpong, sa-
aat.n@sci.kmutnb.ac.th

Academic editor
Haider Mahmood

Additional Information and
Declarations can be found on
page 23

DOI 10.7717/peerj.14194

© Copyright
2022 Maneerat et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

during the transition from the winter to the summer season lead to increased fine particulate matter (PM) levels. In 2020, high PM_{2.5} (PM \leq 2.5 μ m) increased the incidences of several ailments (respiratory illness, allergic reactions in the eyes and nasal passages, *etc.*) in almost all of the regions in Thailand, especially in the northern region ([Tanraksa & Kendall, 2020](#)). Spikes in PM_{2.5} levels in the northern and northeastern regions of Thailand occur during the transition from the winter to the summer season (January–March) ([Wipatayotin & Tangprasert, 2021](#)). These reasons have led to our interest in estimating wind speeds to provide essential information on current PM_{2.5} levels based on historic data. The hourly wind speed data for Phitsanulok and Phayao provinces located in the northern region follow the assumptions for a three-parameter lognormal (TPLN) distribution containing zero values indicating no wind. By way of comparison, estimating the ratio of the median wind speeds in two areas is used as a starting point in this study. The median, a measure of central tendency, is the central value of a dataset (the midpoint of a distribution). Moreover, it is more efficacious to use the median than the mean in analyses when the distribution of data is skewed. In addition, the ratio of the medians of two datasets can be used to measure the difference between them.

A lognormal distribution is used to represent right-skewed data when the threshold parameter (the lower bound of the data) is equal to zero. The TPLN distribution first introduced by [Aitchison & Brown \(1963\)](#) is suitable for highly right-skewed data that do not fit a lognormal distribution because the threshold parameter is greater than zero. It has been used in hydrology ([Borges, Lettenmaier & Bates, 1975](#); [Charbeneau, 1978](#)) and for the analysis of flood frequency ([Singh & Rajagopal, 1986](#); [Singh & Singh, 1987](#)). In this study, zero values are included among both simulated and wind speed data that follow a TPLN distribution (*i.e.*, a TPLN distribution containing zero values).

One of statistical inference methods is the parameter estimations, including the point and interval estimations with the best-known example of the latter being a confidence interval (CI). According to [Casella & Berger \(2002\)](#), the CI is a range of numbers containing the parameter of interest with the desirable level of confidence which is better than a point estimator. For this reason, the CI is focused in this study. The point parameter estimations of the TPLN distribution have been formulated and discussed by various authors. [Cohen & Whitten \(1980\)](#) modified the local maximum likelihood and the moment estimators for the mean, variance and threshold parameters by using the first, second and third order statistics. Next, the moment estimation has been developed by replacing the third moment in a function of the first order statistics, as previously described in [Cohen, Whitten & Ding \(1985\)](#). Later, [Singh, Cruise & Ma \(1990\)](#) compared the five methods, including the regular method of moments, the modified method of moment ([Cohen & Whitten, 1980](#)), the regular maximum likelihood estimate (MLE), the modified MLE, and entropy to estimate the parameters and the quantiles of the TPLN *via* Monte Carlo simulation.

In particular, some researchers have formulated methods for estimating the CIs for the parameters of the TPLN. [Royston \(1992\)](#) used the zero skewness method to estimate the threshold parameter and its certain functions, motivated by [Griffiths \(1980\)](#). After that, [Pang et al. \(2005\)](#) presented the Bayesian estimation using Markov chain Monte Carlo to approximate the coefficient of variation of three-parameter Weibull, lognormal and

gamma distributions. Later, *Basak, Basak & Balakrishnan (2009)* made use of numerical methods (Newton–Raphson and EM algorithms) based on progressively Type-II censored samples from the TPLN for assessing the local and modified MLEs for the mean, variance and threshold parameters as well as the CIs for the threshold by using Monte Carlo simulations. *Chen & Miao (2012)* conducted the order statistics to construct the exact CIs and the exact upper CIs for the threshold parameter. Recently, *Maneerat, Niwitpong & Nakjai (2022b)* formulated the CIs for the median of the TPLN distribution based on bootstrapping, normal approximation (NA) and the generalized pivotal quantity. *Maneerat, Nakjai & Niwitpong (2022a)* also proposed Bayesian confidence intervals based on different noninformative priors for the delta-lognormal mean.

However, the CIs for the parameters of a TPLN distribution, especially when zero values are included along with the nonzero values, have not yet been formulated. Therefore, we herein propose fiducial, NA, and Bayesian methods for constructing CIs for the ratio of the medians of TPLN distributions containing zero values. We conducted a Monte Carlo study to assess their performances in terms of their coverage percentages (CPs), lower and upper error probabilities (LEP and UEP, respectively), and expected lengths. These methods were applied to estimate hourly wind speeds during the transition from the winter to the summer season in northern Thailand. This information could be used to help design green corridors and implement other policies to reduce PM levels in northern Thailand in the future.

THREE-PARAMETER LOGNORMAL DISTRIBUTIONS CONTAINING ZERO VALUES AND THE PARAMETER OF INTEREST

Let $X_i = (X_{i1}, X_{i2}, \dots, X_{in_i})$, $i = 1, 2$ be random samples drawn from a TPLN distribution containing zero values, denoted as $TPLNZ(\gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i)$ where γ_i is the threshold parameter, μ_{X_i} is the scale parameter (the mean of X_i), $\sigma_{X_i}^2$ is the shape parameter (the variance of X_i) and ρ_i is the proportion of nonzero values. For $X_i > 0$, random variable $Y_i = \ln(X_i - \gamma_i)$ is normally distributed as $N(\mu_{X_i}, \sigma_{X_i}^2)$ if X_i is a lognormal random variable (*Cohen & Whitten, 1980*). Thus, the probability density function (pdf) of X_i is given by

$$h(X_i; \gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i) = (1 - \rho_i) + \frac{\rho_i (X_i - \gamma_i)^{-1}}{(2\pi \sigma_{X_i}^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma_{X_i}^2} [\ln(X_i - \gamma_i) - \mu_{X_i}]^2 \right\}, \quad (1)$$

where $\gamma_i < X_i < \infty$, which are bounded as follows: $\gamma_i \geq 0$, $0 < \mu_{X_i} < \infty$, $\sigma_{X_i}^2 > 0$ and $0 < \rho_i < 1$. Otherwise, for $X_i = 0$, the pdf of W becomes $h(X_i; \gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i) = 1 - \rho_i$. The likelihood and log-likelihood functions of $\omega = (\gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i)$ are respectively given by

$$L(\omega|X) = (1 - \rho_i)^{n_{i0}} \rho_i^{n_{i1}} (2\pi \sigma_{X_i}^2)^{n_{i1}/2} \prod_{j=1}^{n_{i1}} \exp \left\{ -\ln(X_{ij} - \gamma_i) - \frac{1}{2\sigma_{X_i}^2} [\ln(X_{ij} - \gamma_i) - \mu_{X_i}]^2 \right\} \quad (2)$$

$$l(\omega|X) = n_{i0} \ln(1 - \rho_i) + n_{i1} \ln \rho_i - \frac{n_{i1}}{2} \ln(2\pi\sigma_{X_i}^2) - \sum_{j:x_{ij}>0} \ln(X_{ij} - \gamma_i) - \frac{1}{2\sigma_{X_i}^2} \sum_{j:x_{ij}>0} [\ln(X_{ij} - \gamma_i) - \mu_{X_i}]^2 \quad (3)$$

where n_{i1} is the number of nonzero values that are binomially distributed with sample size $n_i = n_{i0} + n_{i1}$; ρ_i is the proportion of nonzero values; n_{i0} is the number of zero values. Eq. (3) is computed from the first derivative about $(\gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i)$. The respective MLEs of $(\gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i)$ are solved by setting their first derivative to zero as follows:

$$\frac{\partial l(\omega|X)}{\partial \gamma_i} = \sum_{j:x_{ij}>0} \frac{1}{X_{ij} - \gamma_i} + \frac{n_{i1}}{\sum_{j:x_{ij}>0} [\ln(X_{ij} - \gamma_i) - \frac{1}{n_{i1}} \sum_{j:x_{ij}>0} \ln(X_{ij} - \gamma_i)]^2} \sum_{j:x_{ij}>0} \left[\frac{1}{X_{ij} - \gamma_i} \left\{ \ln(X_{ij} - \gamma_i) - \frac{1}{n_{i1}} \sum_{j:x_{ij}>0} \ln(X_{ij} - \gamma_i) \right\} \right] = 0 \quad (4)$$

$$\frac{\partial l(\omega|X)}{\partial \mu_{X_i}} = \sigma_{X_i}^{-2} \sum_{j:x_{ij}>0} \ln(X_{ij} - \gamma_i) = 0 \quad (5)$$

$$\frac{\partial l(\omega|X)}{\partial \sigma_{X_i}^2} = -(2\sigma_{X_i}^2)^{-1} n_{i1} + [2\sigma_{X_i}^4]^{-1} \sum_{j:x_{ij}>0} [\ln(X_{ij} - \gamma_i) - \mu_{X_i}]^2 = 0 \quad (6)$$

$$\frac{\partial l(\omega|X)}{\partial \rho_i} = n_{i0}(1 - \rho)^{-1} - n_{i1}\rho^{-1} = 0. \quad (7)$$

It is difficult to determine the explicit form of $\hat{\gamma}_i$ from the Eq. (4), so the MLE of γ_i can be obtained by maximizing Eq. (4) using the Adam algorithm. It is a replacement optimization algorithm for stochastic gradient descent based on adaptive moment estimation that is mainly used in neural networks and other machine learning algorithms (Ruder, 2016). It provides a way of computing adaptive learning rates for specific parameters. Let β_1 and β_2 be the initial time step and decay rate, respectively. Adam performs well in practice when $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (Singh, 2020). By optimizing γ_i , the decaying averages and the past squared gradients are the estimates of the first and second moments of the gradients as follows:

$$(m_t, v_t) = (\beta_1 m_{t-1} + [1 - \beta_1] g_t, \beta_2 v_{t-1} + [1 - \beta_2] g_t^2) \quad (8)$$

where g_t denotes the first derivative of the target function (4) at time $t = 0$. Note that (m_t, v_t) are initialized as vectors of 0's, and the Adam algorithm operates until their gradients approach zero. The bias-corrected first- and second-moment estimates are used to update parameter γ_i as follows:

$$(\hat{m}_t, \hat{v}_t) = \left(\frac{m_t}{1 - \beta_1^t}, \frac{v_t}{1 - \beta_2^t} \right). \quad (9)$$

Finally, the parameter γ_i is updated by Adam update rule, which is defined as

$$\gamma_{i,t+1} = \gamma_{i,t} - \delta \left(\sqrt{\hat{v}_t} + \epsilon \right)^{-1} \hat{m}_t \quad (10)$$

where δ and ϵ denote the step size and the learning rate, respectively. ϵ is fixed at 10^{-8} for a sufficient learning rate using the Adam optimization algorithm, as can be seen in Algorithm 1.

Algorithm 1: Adam optimization algorithm

Input: $l, \gamma_{i,t}, \delta, \epsilon, \beta_1, \beta_2$

Output: $\gamma_{i,t}$

```

1 while  $\gamma_{i,t}$  not converged do
2    $t = t + 1$ 
3    $g_t = \nabla_{\gamma_{i,t}} l_t(\gamma_{i,t-1})$ 
4    $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ 
5    $v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ 
6    $\hat{m}_t = m_t / (1 - \beta_1^t)$ 
7    $\hat{v}_t = v_t / (1 - \beta_2^t)$ 
8    $\gamma_{i,t+1} = \gamma_{i,t} - \delta(\sqrt{\hat{v}_t} + \epsilon)^{-1} \hat{m}_t$ 

```

After obtaining $\hat{\gamma}_i$ by using Adam, one can compute the MLEs of $(\mu_{X_i}, \sigma_{X_i}^2, \rho_i)$ as follows:

$$(\hat{\mu}_{X_i}, \hat{\sigma}_{X_i, mle}^2, \hat{\rho}_i) = \left(n_{i1}^{-1} \sum_{j: x_{ij} > 0} \ln(X_{ij} - \hat{\gamma}_i), n_{i1}^{-1} \sum_{j: x_{ij} > 0} [\ln(X_{ij} - \hat{\gamma}_i) - \hat{\mu}_{X_i}]^2, \frac{n_{i1}}{n_i} \right) \quad (11)$$

Since $\eta_i = \rho_i[\gamma_i + \exp(\mu_{X_i})]$ provides the medians of X_i , the ratio of the medians of X_i (the parameter of interest in this study) is given by

$$\omega = \eta_1 / \eta_2. \quad (12)$$

By substituting MLEs $(\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\sigma}_{X_i, mle}^2, \hat{\rho}_i)$, estimate $\hat{\eta}_i = \hat{\rho}_i[\hat{\gamma}_i + \exp(\hat{\mu}_{X_i})]$ becomes the MLE of η_i . The concepts are elaborated in the methods for constructing CIs for ω in the next section.

METHODS

Here, we present constructing the CIs for the ratio of the medians in the TPLN models containing zero observations based on the fiducial methods (the fiducial generalized pivotal quantity (fiducial GPQ)) and the method of variance estimates recovery-based fiducial generalized pivotal quantity (MOVER-fiducial GPQ), NA (*McKean & Schrader, 1984*) and the Bayesian methods.

The fiducial method

Hannig, Iyer & Patterson (2006) described the fiducial method based on the fiducial GPQ, a GPQ subclass introduced by *Weerahandi (1993)*. Fiducial techniques have been used in constructing CIs in several research studies (*Kharrati-Kopaei, Malekzadeh & Sadooghi-Alvandi, 2013; Li, Zhou & Tian, 2013; Maneerat, Niwitpong & Niwitpong, 2020*). Let $Y = (Y_1, Y_2, \dots, Y_n)$ be a random variable with probability density function $f_Y(y; \lambda, \theta)$,

where λ and θ are vectors of the parameter of interest and the nuisance parameter, respectively. Moreover, let $y = (y_1, y_2, \dots, y_n)$ be the observed values of Y , and assume that fiducial GPQ $T(Y; y, \lambda, \theta)$ is only a function of λ . This method is especially associated with the fiducial inference proposed by [Fisher \(1935\)](#). The fiducial GPQ CI depends on the fiducial GPQ defined by [Hannig, Iyer & Patterson \(2006\)](#) in Definition 1. Recently, [Chankham, Niwitpong & Niwitpong \(2022\)](#) recommended the fiducial GPQ-based CI for estimating the coefficient of variation of an inverse gaussian distribution when the sample size was small. Similarly, the performance of a fiducial GPQ CI in terms of expected length was the shortest when used to estimate the common coefficient of variation of delta-lognormal distributions by [Yosboonruang, Niwitpong & Niwitpong \(2022\)](#).

Definition 1.

A GPQ $T(Y; y, \lambda, \theta)$ for a parameter λ is called a *fiducial generalized pivotal quantity* (FGPQ) if it satisfies the following conditions: (FGPQ1) Given $Y = y$, the $T(Y; y, \lambda, \theta)$ distribution is free of all parameters. (FGPQ2) For every $y \in R^+$, the observed pivotal $T(Y; y, \lambda, \theta) = \lambda$.

If the $T(Y; y, \lambda, \theta)$ satisfies the conditions (FGPQ1) and (FGPQ2), then it is possible to construct the $100(1 - \varphi)\%$ fiducial GPQ-based CI for λ is $[T_\lambda(\varphi/2), T_\lambda(1 - \varphi/2)]$; $T_\lambda(\varphi)$ is the φ^{th} percentile of $T(Y; y, \lambda, \theta)$.

The fiducial GPQ CI

The fiducial GPQ CI can be constructed based on the fiducial GPQ conditions in Definition 1. Recall that $(\gamma_i, \mu_{X_i}, \sigma_{X_i}^2, \rho_i)$ are the parameters controlling the behavior of a TPLN containing zero observations. Motivated by the possible value $\gamma_i < X_{i(1)}$, we proposed the fiducial GPQ of γ_i based on a continuously uniform distribution and the fiducial GPQ of ρ_i based on a beta distribution, respectively, as follows:

$$T_{\gamma_i} \sim \text{Uniform}(\min = 0, \max = X_{i(1)}) \quad (13)$$

$$T_{1-\rho_i} \sim \text{Beta}(n_{i0} + 1, n_{i1} + 1) \quad (14)$$

where $X_{i(1)}$ denotes the minimum value of X_i and n_{i1} is the sample size of nonzero values $n_i = n_{i1} + n_{i0}$. Furthermore, the fiducial GPQ of μ_{X_i} can be obtained by using the concepts of [Krishnamoorthy & Mathew \(2003\)](#) as

$$T_{\mu_{X_i}} = \hat{\mu}_{X_i} - W_i \sqrt{T_{\sigma_{X_i}^2} / n_{i1}} \quad (15)$$

where $T_{\sigma_{X_i}^2} = (n_{i1} - 1)\hat{\sigma}_{X_i}^2 / V_i$ is the fiducial GPQ of $\sigma_{X_i}^2$. Moreover, estimator $\hat{\sigma}_{X_i}^2 = (n_{i1} - 1)^{-1} \sum_{j: x_{ij} > 0} \ln(X_{ij} - \hat{\gamma}_i)$. Note that the random variables W_i and V_i are independently draw from a standard normal distribution and a chi-square distribution with $n_{i1} - 1$ degrees of freedom, respectively. Since the three fiducial pivots $(T_{\gamma_i}, T_{\rho_i}, T_{\mu_{X_i}})$ satisfy the fiducial GPQ properties and $T_{\eta_i} = T_{1-\rho_i}[T_{\gamma_i} + \exp(T_{\mu_{X_i}})]$ is the fiducial GPQ of η_i , then we can obtain the the $100(1 - \varphi)\%$ fiducial GPQ CI for ω as follows:

$$[l_\omega^F, u_\omega^F] = [T_\omega(\varphi/2), T_\omega(1 - \varphi/2)] \quad (16)$$

where $T_\omega = T_{\eta_1}/T_{\eta_2}$, and $T_\omega(\varphi)$ denotes the φ^{th} percentile of T_ω . The steps for calculating the CP of fiducial GPQ CI (CP_F) for ω can be carried out by using Algorithm 2.

Algorithm 2: Fiducial GPQ CI

Input: $\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\sigma}_{X_i, mle}^2, \hat{\rho}_i$

Output: CP_F

- 1 **for** k **in** $1:M$ **do**
 - 2 Generate $T_{\gamma_i}, T_{1-\rho_i}$ and $T_{\mu_{X_i}}$
 - 3 Generate T_{η_1} and T_{η_1}
 - 4 Compute T_{ω_i}
 - 5 Compute $[l_\omega^F, u_\omega^F]$
 - 6 if $l_\omega^F < \omega < u_\omega^F$, then $p_k^F = 1$; else $p_k^F = 0$
 - 7 Compute $CP_F = \sum_{i=1}^M p_k^F / M$
-

The MOVER-fiducial GPQ CI

MOVER is a well-known method for estimating the CI of the parameter of interest ([Donner & Zou, 2012](#); [Harvey & Van der Merwe, 2012](#); [Hasan & Krishnamoorthy, 2017](#); [Maneerat & Niwitpong, 2020](#); [Maneerat, Niwitpong & Niwitpong, 2021](#); [Zhang et al., 2021](#); [Maneerat, Nakjai & Niwitpong, 2022a](#)). Moreover, it can produce an explicit form of the CI that is easy to compute. For these reasons, we derived the MOVER-fiducial GPQ CI for ω as follows:

The fiducial GPQ CIs for $(\gamma_i, \mu_{X_i}, \rho_i)$ can respectively be written as

$$[l_{\gamma_i}, u_{\gamma_i}] = [T_{\gamma_i}(\varphi/2), T_{\gamma_i}(\varphi/2)] \quad (17)$$

$$[l_{\mu_{X_i}}, u_{\mu_{X_i}}] = [T_{\mu_{X_i}}(\varphi/2), T_{\mu_{X_i}}(\varphi/2)] \quad (18)$$

$$[l_{\rho_i}, u_{\rho_i}] = [T_{\rho_i}(\varphi/2), T_{\rho_i}(\varphi/2)]. \quad (19)$$

These intervals can be formulated the MOVER CI by using the concept of [Donner & Zou \(2012\)](#) such that the $100(1 - \varphi)\%$ MOVER-fiducial CI for $\ln \eta_i$ becomes

$$[l_{\ln \eta_i}, u_{\ln \eta_i}] = \left[\ln \hat{\eta}_i - \sqrt{\{\ln(\hat{\rho}_i) - l_{\gamma_i}\}^2 + \{\ln[\hat{\gamma}_i + \exp(\hat{\mu}_{X_i})] - l_{\ln[\gamma_i + \exp(\mu_{X_i})]}\}^2}, \ln \hat{\eta}_i + \sqrt{\{\ln(\hat{\rho}_i) - u_{\gamma_i}\}^2 + \{\ln[\hat{\gamma}_i + \exp(\hat{\mu}_{X_i})] - u_{\ln[\gamma_i + \exp(\mu_{X_i})]}\}^2} \right] \quad (20)$$

where

$$l_{\ln[\gamma_i + \exp(\mu_{X_i})]} = \ln \left\{ \hat{\gamma}_i + \exp(\hat{\mu}_{X_i}) - \sqrt{\{\hat{\gamma}_i - l_{\gamma_i}\}^2 + \{\exp(\hat{\mu}_{X_i}) - \exp(l_{\gamma_i})\}^2} \right\} \quad (21)$$

$$u_{\ln[\gamma_i + \exp(\mu_{X_i})]} = \ln \left\{ \hat{\gamma}_i + \exp(\hat{\mu}_{X_i}) + \sqrt{\{\hat{\gamma}_i - l_{\gamma_i}\}^2 + \{\exp(\hat{\mu}_{X_i}) - \exp(l_{\gamma_i})\}^2} \right\}, \quad (22)$$

thereby providing the $100(1 - \varphi)\%$ CI-based the MOVER-fiducial GPQ CI for ω as

$$[l_{\omega}^M, u_{\omega}^M] = \exp \left[(\ln \hat{\eta}_1 - \ln \hat{\eta}_2) - \sqrt{[\ln \hat{\eta}_1 - l_{\ln \hat{\eta}_1}]^2 + [\ln \hat{\eta}_2 - l_{\ln \hat{\eta}_2}]^2}, (\ln \hat{\eta}_1 - \ln \hat{\eta}_2) + \sqrt{[\ln \hat{\eta}_1 - l_{\ln \hat{\eta}_1}]^2 + [\ln \hat{\eta}_2 - l_{\ln \hat{\eta}_2}]^2} \right]. \quad (23)$$

Algorithm 3 presents the computational steps for calculating the CP of the MOVER-fiducial GPQ CI (CP_M) for ω .

Algorithm 3: MOVER-fiducial GPQ CI

Input: $\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\rho}_i$

Output: CP_M

- 1 **for** k in $1:M$ **do**
 - 2 Generate T_{γ_i} to compute $[l_{\gamma_i}, u_{\gamma_i}]$
 - 3 Generate $T_{\mu_{X_i}}$ to compute $[l_{\mu_{X_i}}, u_{\mu_{X_i}}]$
 - 4 Generate T_{ρ_i} to compute $[l_{\rho_i}, u_{\rho_i}]$
 - 5 Compute $[l_{\ln[\gamma_i + \exp(\mu_{X_i})]}, u_{\ln[\gamma_i + \exp(\mu_{X_i})]}]$
 - 6 Compute $[l_{\ln \eta_i}, u_{\ln \eta_i}]$
 - 7 Compute $[l_{\omega}^M, u_{\omega}^M]$
 - 8 if $l_{\omega}^M < \omega < u_{\omega}^M$, then $p_k^M = 1$; else $p_k^M = 0$
 - 9 Compute $CP_M = \sum_{k=1}^M p_k^M / M$
-

The NA method

According to probability theory, the concept behind this method is the assumption that the approximate distributions of all of the samples approach a normal distribution pattern if the sample size is sufficiently large. This idea is integrated with the central limit theorem, in which the distribution of a given sample mean is approximated as a normal pattern if the sample size is sufficiently large under the assumption that all of the samples are similar to each other regardless of the shape of the population distribution. Recently, [Maneerat, Niwitpong & Nakjai \(2022b\)](#) proposed an NA-based CI for the median of a TPLN distribution, which performed well for a large sample size. Thus, we also considered the NA method.

Given a set of observations, threshold γ_i can be estimated by using the Adam algorithm to find the MLEs of the mean and variance ($\hat{\mu}_{X_i}, \hat{\sigma}_{X_i}^2$). Here, the medians of TPLN models with zero observations can be log-transformed to become

$$\ln \eta_i = \ln \rho_i + \ln[\gamma_i + \exp(\mu_{X_i})] \quad (24)$$

which can be approximated by using $(\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\rho}_i)$ to give $\ln \hat{\eta}_i = \ln \hat{\rho}_i + \ln[\hat{\gamma}_i + \exp(\hat{\mu}_{X_i})]$. Using the delta method, the variance of $\ln \rho_i$ can be derived as

$$V_{\ln \hat{\rho}_i} = \frac{1 - \rho_i}{n \rho_i}. \quad (25)$$

Likewise, *McKean & Schrader (1984)* estimated the variance of the median as a distribution-free estimate defined as follows:

$$V_{\hat{\eta}_i}^{MS} = \left[\frac{X_{i(n_{i1}-c+1)} + X_{i(c)}}{2(1.96)} \right]^2 \quad (26)$$

where $X_{i(r)}$ denotes the r^{th} order statistic in a random sample drawn from a TPLN model with zero observations of size n_{i1} , and $c = [(n_{i1} + 1)/2] - 1.96(\sqrt{n_{i1}/4})$. Later, *Hettmansperger & Sheather (1986)* claimed that V^{MS} is a consistent estimator of the variance of the median. Similarly, the variance of $\ln \hat{\eta}_i$ can be derived by applying the delta method as follows:

$$V_{\ln \hat{\eta}_i} \cong \frac{V_{\hat{\eta}_i}^{MS}}{[E_{\hat{\eta}_i}]^2} \quad (27)$$

where $E_{\hat{\eta}_i} = \hat{\gamma}_i + \exp[\hat{\mu}_{X_i} + \hat{\sigma}_{X_i}^2(2n_{i1})^{-1}]$ is the expectation of $\hat{\eta}_i$. Estimated variance $\widehat{V}_{\ln \hat{\eta}_i}$ is obtainable by replacing $(\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\sigma}_{X_i}^2, \hat{\rho}_i)$ from the sample. Thus, by applying the central limit theorem, the random variable W_i can be defined as

$$W_i = \frac{\ln \hat{\eta}_i - \ln \eta_i}{\sqrt{\widehat{V}_{\ln \hat{\eta}_i}}} \quad (28)$$

which approaches a standard normal distribution as $n \rightarrow \infty$. Subsequently, the $100(1 - \varphi)\%$ NA-based CI for ω can be written as

$$[l_{\omega}^N, u_{\omega}^N] = \exp \left[(\ln \hat{\eta}_1 - \ln \hat{\eta}_2) \mp W_{1-\varphi/2} \sqrt{\widehat{V}_{\ln \hat{\eta}_1} + \widehat{V}_{\ln \hat{\eta}_2}} \right] \quad (29)$$

where W_{φ} denotes the φ^{th} percentile of a standard normal distribution. Algorithm 4 was used to compute the CP of the NA (CP_N).

Algorithm 4: NA CI

Input: $\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\rho}_i$

Output: CP_N

- 1 **for** k in $1:M$ **do**
 - 2 Compute $V_{\ln \hat{\rho}_i}$ and $V_{\hat{\eta}_i}^{MS}$
 - 3 Compute $V_{\hat{\eta}_i}$
 - 4 Generate W_i
 - 5 Compute $[l_{\omega}^N, u_{\omega}^N]$
 - 6 if $l_{\omega}^N < \omega < u_{\omega}^N$, then $p_k^N = 1$; else $p_k^N = 0$
 - 7 **Compute** $CP_N = \sum_{k=1}^M p_k^N / M$
-

The Bayesian method

Bayesian methods are based on treating probability as beliefs rather than frequencies. Given unknown parameter θ , a prior distribution $p(\theta)$ represents the subjective belief as a subjective distribution formulated before the data are seen. The posterior distribution is

obtained from a prior that is updated with the likelihood function (or sample information) by using Bayes' rule (Casella & Berger, 2002). Importantly, the posterior distribution is considered to be a random quantity and can be used to make a statement about θ . For example, the point and interval estimates of θ can be computed by using its posterior. Equal-tailed Bayesian intervals based on the Jeffreys' and uniform priors based on the posterior densities of the zero proportion and the variance have been shown to perform well in certain scenarios (Yosboonruang, Niwitpong & Niwitpong, 2022).

Here, the Bayesian CI for ω is formulated based on the Bayesian method. First, we define an informative prior for our objective assumption depending on the amount of information available in the data as

$$P(\gamma_i, \mu_{X_i}, \log \sigma_{X_i}, \rho_i) = \text{constant} * (1 - \rho_i)^{\alpha_i - 1} \rho_i^{\beta_i - 1} \Gamma(\alpha_i + \beta_i) [\Gamma(\alpha_i) \Gamma(\beta_i)]^{-1}. \quad (30)$$

The posterior densities of (γ_i, ρ_i) are obtained by obtaining Eq. (30) with the likelihood function (2) as

$$\begin{aligned} f_{\gamma_i} = P(\gamma_i | X, \hat{\mu}_{X_i}, \hat{\sigma}_{X_i}^2) &= \left[\text{constant} * (1 - \rho_i)^{\alpha_i - 1} \rho_i^{\beta_i - 1} \Gamma(\alpha_i + \beta_i) [\Gamma(\alpha_i) \Gamma(\beta_i)]^{-1} \right] \\ &* \left[(1 - \rho_i)^{n_{i0}} \rho_i^{n_{i1}} (2\pi \sigma_{X_i}^2)^{-n_{i1}/2} \exp \left\{ - \sum_{j: X_{ij} > 0} \ln(X_{ij} - \gamma_i) - \frac{1}{2\sigma_{X_i}^2} \sum_{j: X_{ij} > 0} [\ln(X_{ij} - \gamma_i) - \mu_{X_i}]^2 \right\} \right] \\ &\propto (\hat{\sigma}_{X_i}^2)^{-n_{i1}/2} \exp \left\{ - \sum_{j: X_{ij} > 0} \ln(X_{ij} - \gamma_i) - \frac{1}{2\hat{\sigma}_{X_i}^2} \sum_{j: X_{ij} > 0} [\ln(X_{ij} - \gamma_i) - \hat{\mu}_{X_i}]^2 \right\} \end{aligned} \quad (32)$$

$$\begin{aligned} f_{\rho_i} = P(\rho_i | X) &= \left[\text{constant} * (1 - \rho_i)^{n_{i0} + \alpha_i - 1} \rho_i^{n_{i1} \beta_i - 1} \Gamma(\alpha_i + \beta_i) [\Gamma(\alpha_i) \Gamma(\beta_i)]^{-1} \right] \\ &* \left[(2\pi \sigma_{X_i}^2)^{n_{i1}/2} \exp \left\{ - \sum_{j: X_{ij} > 0} \ln(X_{ij} - \gamma_i) - \frac{1}{2\sigma_{X_i}^2} \sum_{j: X_{ij} > 0} [\ln(X_{ij} - \gamma_i) - \mu_{X_i}]^2 \right\} \right] \end{aligned} \quad (33)$$

$$\propto (1 - \rho_i)^{n_{i0} + \alpha_i - 1} \rho_i^{n_{i1} \beta_i - 1}. \quad (34)$$

Next, we apply NA to the posterior distribution of $\kappa_i = (\mu_{X_i}, \log \sigma_{X_i})$; the logarithm of the posterior density is approximated by using a quadratic function of κ_i . The second derivatives of the log-posterior density are needed for constructing the approximation. From Eq. (3), the log-likelihood can be expressed as

$$\begin{aligned} l(\kappa_i | X) &= n_{i0} \ln(1 - \rho_i) + n_{i1} \ln \rho_i - \frac{n_{i1}}{2} \ln(2\pi) + n_{i1} \ln \sigma_{X_i} \\ &- \sum_{j: X_{ij} > 0} \ln(X_{ij} - \gamma_i) - \frac{1}{2\sigma_{X_i}^2} [(n_{i1} - 1) \hat{\sigma}_{X_i}^2 + n_{i1} (\hat{\mu}_{X_i} - \mu_{X_i})^2] \propto \text{constant}. \\ &+ n_{i1} \ln \sigma_{X_i} - \frac{1}{2 \exp(2 \ln \sigma_{X_i})} [(n_{i1} - 1) \hat{\sigma}_{X_i}^2 + n_{i1} (\hat{\mu}_{X_i} - \mu_{X_i})^2]. \end{aligned} \quad (35)$$

After that the first and second derivatives of $(\mu_{X_i}, \log \sigma_{X_i})$ respectively become

$$\left(\frac{\partial l(\kappa_i | X)}{\partial \mu_{X_i}}, \frac{\partial^2 l(\kappa_i | X)}{\partial \mu_{X_i}^2} \right) = \left(\frac{n_{i1} (\hat{\mu}_{X_i} - \mu_{X_i})}{\sigma_{X_i}^2}, -\frac{n_{i1}}{\sigma_{X_i}^2} \right) \quad (36)$$

$$\left(\frac{\partial l(\kappa_i|X)}{\partial \ln \sigma_{X_i}}, \frac{\partial^2 l(\kappa_i|X)}{\partial [\ln \sigma_{X_i}]^2} \right) = \left(-n_{i1} + \sigma_{X_i}^{-2} [(n_{i1} - 1)\hat{\sigma}_{X_i}^2 + n_{i1}(\hat{\mu}_{X_i} - \mu_{X_i})^2], \right. \\ \left. -2\sigma_{X_i}^{-1} [(n_{i1} - 1)\hat{\sigma}_{X_i}^2 + n_{i1}(\hat{\mu}_{X_i} - \mu_{X_i})^2] \right). \quad (37)$$

The point estimates of $(\mu_{X_i}, \log \sigma_{X_i})$ are derived after setting their first derivatives to zero. Meanwhile, the variances of their estimates are respectively obtained by using an inverse Fisher information matrix as follows:

$$\begin{pmatrix} \hat{\mu}_{X_i} \\ \ln \hat{\sigma}_{X_i} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_{X_i} \\ \ln \sqrt{(n_{i1} - 1)n_{i1}^{-1}\hat{\sigma}_{X_i}^2} \end{pmatrix} \quad (38)$$

$$\begin{pmatrix} V_{\hat{\mu}_{X_i}} \\ V_{\ln \hat{\sigma}_{X_i}} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{X_i}^2/n_{i1} \\ 2n_{i1} \end{pmatrix}. \quad (39)$$

After using the Jacobian to back-transform from $\ln \hat{\sigma}_{X_i}$ to $\sigma_{X_i}^2$, the posterior densities of $(\mu_{X_i}, \sigma_{X_i}^2)$ are respectively approximated as normal distribution as follows:

$$f_{\mu_{X_i}} = p(\mu_{X_i}|X, \hat{\sigma}_{X_i}^2, \hat{\gamma}_i) = \frac{1}{(2\pi\hat{\sigma}_{X_i}^2/n_{i1})^{1/2}} \exp \left\{ -\frac{n_{i1}}{2\hat{\sigma}_{X_i}^2} [\ln(X_{ij} - \hat{\gamma}_i) - \hat{\mu}_{X_i}]^2 \right\} \quad (40)$$

$$f_{\sigma_{X_i}^2} = p(\sigma_{X_i}^2|X) = \left(\frac{n_{i1} + 2}{4\pi\hat{\sigma}_{X_i}^4} \right)^{1/2} \exp \left\{ -\frac{n_{i1} + 2}{4\hat{\sigma}_{X_i}^4} [\ln(X_{ij} - \hat{\gamma}_i) - \hat{\sigma}_{X_i}^2]^2 \right\} \quad (41)$$

where $\hat{\sigma}_{X_i}^2 = n_{i1}\hat{\sigma}_{X_i}^2/(n_{i1} + 2)$. Therefore, the posterior density of ω becomes

$$f_{\omega} = \frac{f_{\rho_1} [f_{\gamma_1} + \exp(f_{\mu_{X_1}})]}{f_{\rho_2} [f_{\gamma_2} + \exp(f_{\mu_{X_2}})]} \quad (42)$$

Finally, the $100(1 - \varphi)\%$ Bayesian-based CI for ω is

$$[l_{\omega}^B, u_{\omega}^B] = [f_{\omega}(\varphi/2), f_{\omega}(1 - \varphi/2)] \quad (43)$$

where $f_{\omega}(\varphi)$ denotes the φ^{th} percentile of f_{ω} . The CP of Bayesian CI (CP_B) can be computed by using Algorithm 5.

Algorithm 5: Bayesian CI

Input: $\hat{\gamma}_i, \hat{\mu}_{X_i}, \hat{\sigma}_{X_i}^2, \hat{\rho}_i, \hat{\sigma}_{X_i}, \alpha_i, \beta_i$

Output: CP_B

- 1 **for** k **in** $1:M$ **do**
 - 2 Generate $f_{\gamma_i}, f_{\rho_i}, f_{\mu_{X_i}}$
 - 3 Compute f_{ω}
 - 4 Compute $[l_{\omega}^B, u_{\omega}^B]$
 - 5 if $l_{\omega}^B < \omega < u_{\omega}^B$, then $p_k^B = 1$; else $p_k^B = 0$
 - 6 **Compute** $CP_B = \sum_{k=1}^M p_k^B / M$
-

THE MONTE CARLO SIMULATION STUDY

The comparative performances of the proposed Bayesian-, the fiducial GPQ- and MOVER-fiducial GPQ-, and NA-based CI were evaluated *via* a Monte Carlo simulation study. The settings for the simulation parameters for the two simulation studies were as follows. For each specified parameter combination, the 95% CIs for the ratio of the medians (ω) of TPLN distributions containing zero values were constructed based on 5000 randomly generated samples. In addition, 2500 Monte Carlo sampling passes were used for each generated sample for the fiducial GPQ-based method. To assess the performances of the methods, their CPs were calculated by using the proportion of 5000 simulated CIs covering ω . LEP and UEP are defined as the proportion of times that ω falls below and above the stimulated CIs, respectively. At the 95% nominal confidence level, the expected lengths (ELs) of the CIs is also needed for deciding which method performs the best. A good performance will produce $CP = 95\%$ and $LEP = UEP = 2.5\%$. Likewise, the comparison between LEP and UEP can be expressed in terms of the relative bias, which is defined as

$$RB = \frac{UEP - LEP}{UEP + LEP}. \quad (44)$$

Thus, a good balance between LEP and UEP will produce a relative bias close to zero. Last, the best-performing method will provide the shortest EL.

In the first simulation study, we chose a small proportion of zeros and variance ($(d_1, d_2) = (10\%, 10\%), (10\%, 30\%)$ and $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1.25$, respectively), the results of which provide insight into the sampling behavior of the CIs (Table 1 and Figs. 1, 2 and 3). It can be seen that although all of the methods generated CPs above or close to the nominal confidence level in almost all of the scenarios, the LEPs, UEPs, and ELs produced by the Bayesian method demonstrated its superiority. Thus, the Bayesian method performed the best in situations with a small proportion of zeros and sample variance except for with a large sample size, with the MOVER-fiducial GPQ method performing the best in that case.

In the next simulation, we were interested in scenarios with a large proportion of zeros and variance ($d_1 = d_2 = (20\%, 40\%), (40\%, 40\%)$ and $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 3$, respectively) (Table 2 and Figs. 4, 5 and 6). Once again, the Bayesian method provided acceptable CPs, as well as better ELs and a better balance between LEP and UEP, than the other methods.

APPLICATION OF THE METHODS TO COMPARE HOURLY WIND SPEED DATA FROM TWO AREAS IN NORTHERN THAILAND

Due to the rapid effects of climate change, agricultural growth, and the social economy, seasonal air pollution from the burning of agricultural waste in preparation for planting, forest fires, and waste disposal during the transition from the winter to the dry season are important factors that influence the environment in northern Thailand (IQAir, 2021). Wind can affect the movement of $PM_{2.5}$ and PM_{10} when its speed is 7.2 km/hr or higher (Liu et al., 2020). When cold air mass moves from China to Thailand, the upper region of Thailand can potentially become very cold toward the end of winter, which reduces

Table 1 Monte Carlo simulation results from the simulation study 1: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1.25$.

Scenarios	(n_1, n_2)	$(\rho_1, \rho_2)\%$	(γ_1, γ_2)	B				F				M				N				Relative bias			
				LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	B	F	M	N
1	(25,25)	(10,10)	(1,1)	0.72	98.30	0.98	163.37	0.18	99.52	0.30	200.56	0.06	99.48	0.46	189.01	0.34	99.18	0.48	211.70	0.15	0.25	0.77	0.79
2			(1,3)	0.66	98.44	0.90	156.22	0.28	99.48	0.24	198.36	0.00	99.58	0.42	186.41	0.28	99.26	0.46	202.52	0.15	-0.08	1.00	1.00
3			(3,5)	0.74	98.44	0.82	142.20	0.24	99.68	0.08	195.00	0.00	99.80	0.20	182.46	0.36	99.28	0.36	182.52	0.05	-0.50	1.00	1.00
4		(10,30)	(1,1)	0.96	98.30	0.74	180.80	0.36	99.42	0.22	222.67	0.18	99.58	0.24	208.90	0.48	99.14	0.38	240.36	0.13	-0.24	0.14	1.00
5			(1,3)	0.90	97.98	1.12	171.38	0.54	99.22	0.24	218.83	0.04	99.46	0.50	204.42	0.44	99.00	0.56	225.81	0.11	-0.38	0.85	1.00
6			(3,5)	0.90	98.02	1.08	157.74	0.28	99.54	0.18	215.19	0.08	99.52	0.40	200.41	0.44	99.08	0.48	206.96	0.09	-0.22	0.67	1.00
7	(25,50)	(10,10)	(1,1)	0.38	96.92	2.70	136.08	0.12	98.84	1.04	167.07	0.12	99.00	0.88	158.35	0.14	98.98	0.88	177.91	0.75	0.79	0.76	0.56
8			(1,3)	0.48	96.26	3.26	132.95	0.12	98.94	0.94	167.09	0.04	98.64	1.32	158.41	0.26	98.42	1.32	173.17	0.74	0.77	0.94	1.00
9			(3,5)	0.28	96.96	2.76	121.35	0.02	99.36	0.62	165.16	0.02	98.96	1.02	156.11	0.14	98.86	1.00	156.03	0.82	0.94	0.96	1.00
10		(10,30)	(1,1)	0.70	96.48	2.82	146.90	0.20	98.60	1.20	179.03	0.12	98.90	0.98	169.60	0.36	98.80	0.84	190.57	0.60	0.71	0.78	1.00
11			(1,3)	0.44	96.86	2.70	142.87	0.18	98.86	0.96	178.49	0.06	98.70	1.24	168.96	0.36	98.70	0.94	183.63	0.72	0.68	0.91	1.00
12			(3,5)	0.54	96.94	2.52	130.98	0.16	99.28	0.56	176.39	0.02	98.96	1.02	166.45	0.32	98.88	0.80	167.46	0.65	0.56	0.96	1.00
13	(50,50)	(10,10)	(1,1)	2.20	95.28	2.52	103.10	0.74	98.34	0.92	125.77	0.38	98.60	1.02	121.00	0.56	98.66	0.78	137.55	0.07	0.11	0.46	0.47
14			(1,3)	2.30	95.34	2.36	99.18	1.12	98.40	0.48	125.60	0.18	98.80	1.02	120.90	0.46	98.90	0.64	131.36	0.01	-0.40	0.70	1.00
15			(3,5)	2.20	95.40	2.40	91.07	0.70	99.10	0.20	126.29	0.00	98.96	1.04	121.84	0.48	99.02	0.50	119.67	0.04	-0.56	1.00	1.00
16		(10,30)	(1,1)	2.32	96.16	1.52	116.99	0.90	98.52	0.58	140.93	0.46	98.80	0.74	135.19	0.64	99.04	0.32	153.10	0.21	-0.22	0.23	1.00
17			(1,3)	2.44	95.88	1.68	111.47	1.14	98.36	0.50	140.01	0.18	98.92	0.90	134.02	0.70	98.72	0.58	145.15	0.18	-0.39	0.67	1.00
18			(3,5)	2.52	95.76	1.72	102.94	0.88	98.86	0.26	139.91	0.04	99.06	0.90	134.30	0.66	98.74	0.60	132.03	0.19	-0.54	0.91	1.00
19	(50,100)	(10,10)	(1,1)	1.28	92.60	6.12	86.81	0.36	97.48	2.16	106.31	0.28	97.58	2.14	102.76	0.42	98.28	1.30	116.03	0.65	0.71	0.77	0.07
20			(1,3)	1.34	92.66	6.00	84.88	0.64	98.04	1.32	107.17	0.12	97.80	2.08	104.00	0.34	98.38	1.28	112.56	0.63	0.35	0.89	1.00
21			(3,5)	1.18	92.76	6.06	77.90	0.28	98.80	0.92	108.69	0.00	97.64	2.36	105.99	0.18	98.52	1.30	102.12	0.67	0.53	1.00	1.00
22		(10,30)	(1,1)	1.56	94.06	4.38	93.63	0.50	97.66	1.84	113.43	0.34	97.88	1.78	109.72	0.38	98.58	1.04	123.78	0.47	0.57	0.68	1.00
23			(1,3)	1.56	94.08	4.36	91.27	0.80	97.98	1.22	114.13	0.18	97.98	1.84	110.65	0.34	98.66	1.00	119.67	0.47	0.21	0.82	1.00
24			(3,5)	1.38	94.16	4.46	84.33	0.44	98.78	0.78	115.57	0.02	98.16	1.82	112.46	0.42	98.52	1.06	109.19	0.53	0.28	0.98	1.00
25	(100,100)	(10,10)	(1,1)	3.94	92.68	3.38	66.72	1.40	97.36	1.24	82.27	0.66	98.10	1.24	80.42	0.96	98.42	0.62	90.22	0.08	-0.06	0.31	-0.05
26			(1,3)	4.22	92.16	3.62	64.20	2.08	97.30	0.62	83.51	0.12	98.12	1.76	82.05	0.74	98.56	0.70	86.14	0.08	-0.54	0.87	1.00
27			(3,5)	4.36	91.82	3.82	59.41	0.82	98.80	0.38	87.00	0.00	98.32	1.68	86.29	0.64	98.54	0.82	78.57	0.07	-0.37	1.00	1.00
28		(10,30)	(1,1)	4.32	92.50	3.18	75.39	1.56	97.30	1.14	91.48	0.68	97.88	1.44	89.20	0.86	98.32	0.82	99.62	0.15	-0.16	0.36	1.00
29			(1,3)	4.64	92.48	2.88	72.36	2.00	97.58	0.42	92.34	0.14	98.58	1.28	90.33	0.58	98.70	0.72	94.97	0.23	-0.65	0.80	1.00

(continued on next page)

Table 1 (continued)

Scenarios	(n_1, n_2)	$(\rho_1, \rho_2)\%$	(γ_1, γ_2)	B				F				M				N				Relative bias			
				LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	B	F	M	N
30			(3,5)	4.78	92.48	2.74	67.33	1.46	98.32	0.22	95.34	0.04	98.74	1.22	94.07	0.76	98.56	0.68	87.22	0.27	-0.74	0.94	1.00
31	(100,200)	(10,10)	(1,1)	1.66	89.92	8.42	56.58	0.54	96.54	2.92	70.38	0.30	96.88	2.82	69.05	0.36	98.42	1.22	76.93	0.67	0.69	0.81	-0.44
32			(1,3)	2.38	90.08	7.54	55.34	1.14	97.62	1.24	72.25	0.18	97.10	2.72	71.61	0.64	98.02	1.34	74.88	0.52	0.04	0.88	1.00
33			(3,5)	2.14	90.02	7.84	51.27	0.62	98.86	0.52	76.67	0.00	97.22	2.78	76.83	0.48	97.98	1.54	68.25	0.57	-0.09	1.00	1.00
34	(10,30)		(1,1)	2.42	91.16	6.42	61.27	0.90	97.04	2.06	75.23	0.34	97.48	2.18	73.82	0.58	98.46	0.96	81.88	0.45	0.39	0.73	1.00
35			(1,3)	2.64	90.74	6.62	59.68	1.28	97.66	1.06	76.81	0.18	97.06	2.76	75.98	0.50	98.16	1.34	79.33	0.43	-0.09	0.88	1.00
36			(3,5)	3.00	90.52	6.48	55.51	0.88	98.46	0.66	80.91	0.04	97.10	2.86	80.90	0.76	97.88	1.36	72.88	0.37	-0.14	0.97	1.00

Notes.

B, the Bayesian CI; F, the fiducial GPQ CI; M, the MOVER-fiducial GPQ CI; N, the NA CI.

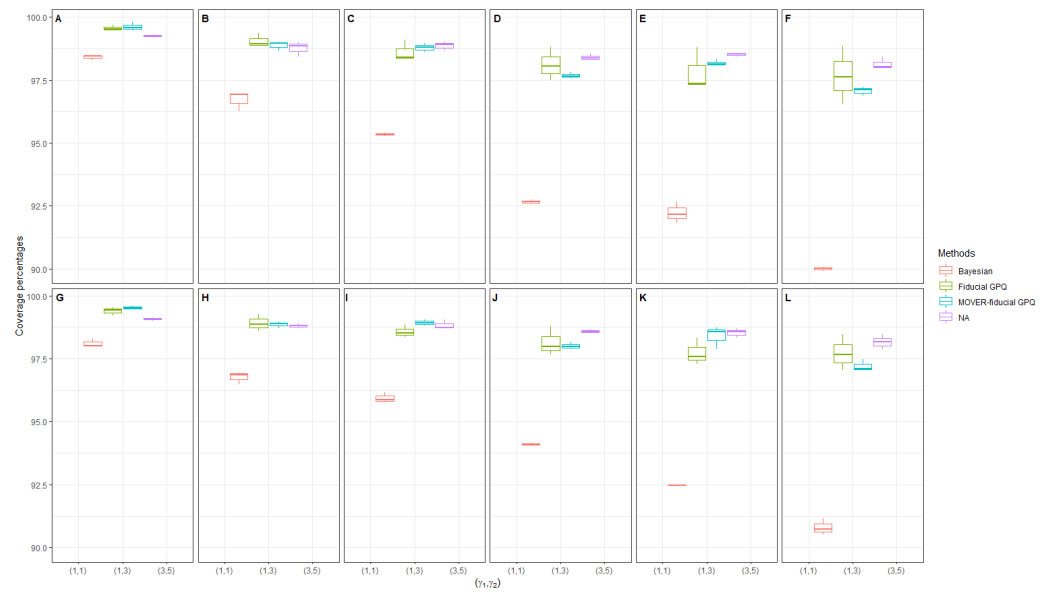


Figure 1 Simulation 1: Box plots of the coverage percentages for 36 scenarios based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1.25$ in the following cases [sample sizes, zero proportions]: (A)[(25,25),(10,10)%],(B)[(25,50),(10,10)%], (C)[(50,50),(10,10)%],(D)[(50,100),(10,10)%], (E)[(100,100),(10,10)%],(F)[(100,200),(10,10)%], (G)[(25,25),(10,30)%],(H)[(25,50),(10,30)%], (I)[(50,50),(10,30)%],(J)[(50,100),(10,30)%], (K)[(100,100),(10,30)%],(L)[(100,200),(10,30)%].

Full-size DOI: 10.7717/peerj.14194/fig-1

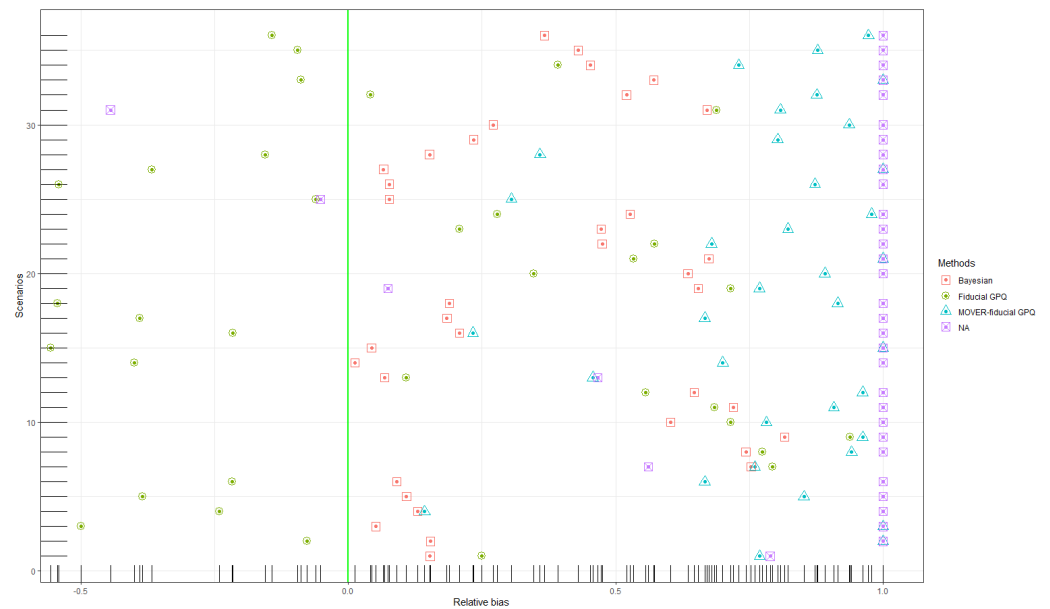


Figure 2 Simulation 1: Scatter plots of the relative bias for Scenarios 1–36 based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1.25$.

Full-size DOI: 10.7717/peerj.14194/fig-2

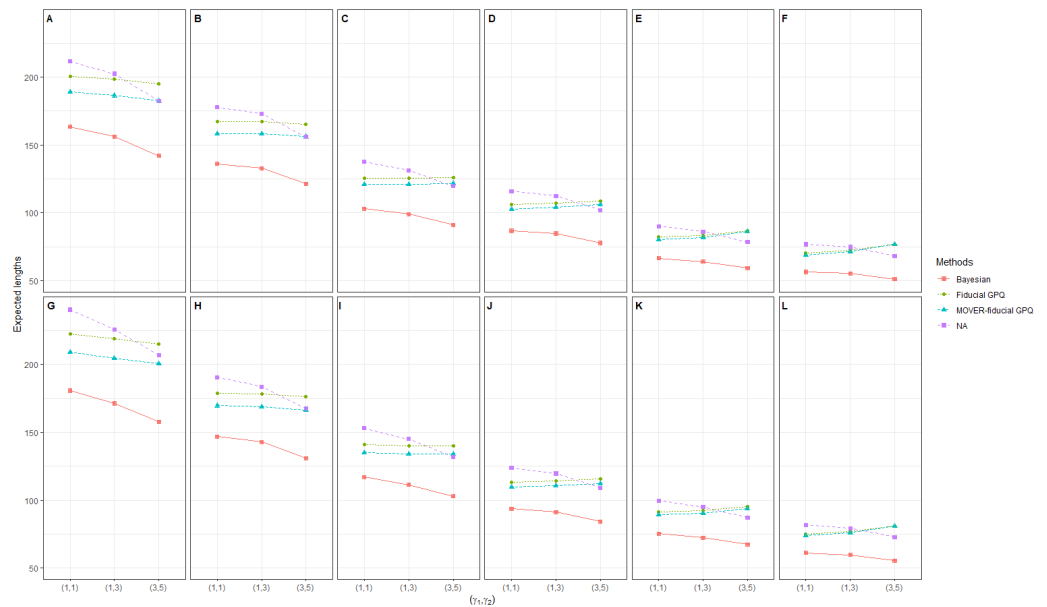


Figure 3 Simulation 1: Line plots of the expected width for 36 scenarios based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1.25$ in the following cases [sample sizes, zero proportions]: (A)[(25,25),(10,10)%],(B)[(25,50),(10,10)%], (C)[(50,50),(10,10)%],(D)[(50,100),(10,10)%], (E)[(100,100),(10,10)%],(F)[(100,200),(10,10)%], (G)[(25,25),(10,30)%],(H)[(25,50),(10,30)%], (I)[(50,50),(10,30)%],(J)[(50,100),(10,30)%], (K)[(100,100),(10,30)%],(L)[(100,200),(10,30)%].

Full-size [DOI: 10.7717/peerj.14194/fig-3](https://doi.org/10.7717/peerj.14194/fig-3)

the northeast monsoon to a calm wind. For this reason, $PM_{2.5}$ levels usually increase during the transition between the winter season to the dry season (Teerasuphaset & Culp, 2020). Phitsanulok is a city in lower northern Thailand about halfway between Chiang Mai and Bangkok where crop and forestland burning is extensive, resulting in extreme $PM_{2.5}$ occurrences (IQAir, 2022), while Phayao is one of the three highest-ranking provinces for $PM_{2.5}$ in the upper northern region (Group, 2021).

We used datasets of the hourly wind speed from Phitsanulok and Phayao (Table 3) recorded in January 2021 to illustrate the efficacies of our proposed methods for formulating CIs for the ratio of the medians of TPLN distributions containing zero values. The data were taken from the Thai Meteorological Department Automatic Weather System (Thai Meteorological Department Automatic Weather System, 2022). Since wind speed observations are always non-negative, they are suitable for fitting to the following distributions: Cauchy, chi-squared, exponential, lognormal, TPLN, logistic, normal, and t-distributions. The Akaike information criterion (AIC) can be used to determine the best-fitting distribution.

We found that the TPLN model was suitable for the wind speed data, as evidenced by its smallest AIC value in Table 4. The basic statistics for the datasets are reported in Table 5. By way of comparison, the ratio of the medians of the TPLN distributions of the hourly wind speed data from Phayao and Phitsanulok is $\hat{\omega} = 1.0217$ for medians $\hat{\eta}_{phayao} = 7.4091$ and $\hat{\eta}_{phitsanulok} = 7.2514$. The 95% CIs and corresponding lengths based on the fiducial,

Table 2 Monte Carlo simulation results from the simulation study 2: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 3$.

Scenarios	(n_1, n_2)	$(\rho_1, \rho_2)\%$	(γ_1, γ_2)	B				F				M				N				Relative bias			
				LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	B	F	M	N
37	(25,25)	(20,40)	(1,1)	1.70	97.30	1.00	259.44	1.28	98.10	0.62	293.83	1.52	97.88	0.60	280.54	0.40	99.30	0.30	444.30	-0.26	-0.35	-0.43	0.89
38			(1,3)	1.36	97.60	1.04	245.93	1.08	98.52	0.40	286.48	0.92	98.70	0.38	269.61	0.46	99.30	0.24	417.03	-0.13	-0.46	-0.42	1.00
39			(3,5)	1.44	97.24	1.32	224.79	1.04	98.40	0.56	276.06	0.46	98.92	0.62	255.06	0.34	99.10	0.56	380.39	-0.04	-0.30	0.15	1.00
40		(40,40)	(1,1)	1.66	97.02	1.32	283.20	1.00	98.28	0.72	323.39	1.38	97.84	0.78	307.55	0.24	99.36	0.40	516.55	-0.11	-0.16	-0.28	1.00
41			(1,3)	0.78	98.06	1.16	270.82	0.42	99.00	0.58	317.18	0.50	98.72	0.78	297.65	0.20	99.52	0.28	486.72	0.20	0.16	0.22	1.00
42			(3,5)	1.48	97.20	1.32	248.95	0.82	98.58	0.60	306.52	0.46	98.88	0.66	282.32	0.36	99.40	0.24	445.82	-0.06	-0.15	0.18	1.00
43	(25,50)	(20,40)	(1,1)	1.22	96.54	2.24	213.10	0.92	97.48	1.60	235.89	1.06	97.38	1.56	227.43	0.44	99.02	0.54	320.58	0.29	0.27	0.19	0.73
44			(1,3)	1.36	95.90	2.74	205.82	1.16	97.12	1.72	232.74	0.92	97.18	1.90	222.39	0.64	98.62	0.74	309.26	0.34	0.19	0.35	1.00
45			(3,5)	1.14	96.56	2.30	188.47	0.82	97.82	1.36	225.40	0.42	98.10	1.48	211.28	0.26	99.02	0.72	283.15	0.34	0.25	0.56	1.00
46		(40,40)	(1,1)	1.00	96.68	2.32	242.04	0.70	97.66	1.64	272.84	0.82	97.60	1.58	260.74	0.40	98.94	0.66	404.46	0.40	0.40	0.32	1.00
47			(1,3)	0.70	96.42	2.88	235.54	0.56	97.68	1.76	270.25	0.48	97.60	1.92	256.41	0.34	98.80	0.86	395.13	0.61	0.52	0.60	1.00
48			(3,5)	1.04	96.58	2.38	214.38	0.72	97.78	1.50	259.15	0.48	97.84	1.68	241.05	0.38	98.86	0.76	360.16	0.39	0.35	0.56	1.00
49	(50,50)	(20,40)	(1,1)	1.96	96.32	1.72	177.64	1.52	97.26	1.22	193.23	1.60	97.12	1.28	187.25	0.90	98.70	0.40	249.14	-0.07	-0.11	-0.11	0.67
50			(1,3)	2.02	96.36	1.62	168.55	2.06	96.98	0.96	189.16	1.28	97.54	1.18	180.85	0.76	98.72	0.52	237.29	-0.11	-0.36	-0.04	1.00
51			(3,5)	1.88	95.98	2.14	155.15	1.64	97.46	0.90	184.51	0.74	98.02	1.24	173.93	0.76	98.64	0.60	218.59	0.06	-0.29	0.25	1.00
52		(40,40)	(1,1)	1.50	97.02	1.48	192.68	1.20	97.58	1.22	209.88	1.20	97.64	1.16	203.27	0.54	98.90	0.56	273.71	-0.01	0.01	-0.02	1.00
53			(1,3)	1.96	96.44	1.60	184.84	2.00	97.12	0.88	206.78	1.26	97.58	1.16	198.06	0.96	98.42	0.62	261.51	-0.10	-0.39	-0.04	1.00
54			(3,5)	1.48	96.80	1.72	169.90	1.28	97.74	0.98	200.86	0.58	98.20	1.22	189.24	0.52	98.74	0.74	241.20	0.08	-0.13	0.36	1.00
55	(50,100)	(20,40)	(1,1)	1.56	96.02	2.42	147.79	1.12	97.06	1.82	159.74	1.24	97.00	1.76	155.22	0.52	98.76	0.72	202.75	0.22	0.24	0.17	0.34
56			(1,3)	1.30	96.16	2.54	142.96	1.26	97.20	1.54	158.18	0.74	97.24	2.02	152.63	0.48	98.98	0.54	196.49	0.32	0.10	0.46	1.00
57			(3,5)	1.36	96.22	2.42	131.20	1.24	97.54	1.22	154.66	0.46	97.88	1.66	147.32	0.42	98.82	0.76	180.39	0.28	-0.01	0.57	1.00
58		(40,40)	(1,1)	1.38	96.32	2.30	165.56	1.06	96.80	2.14	179.74	1.10	96.86	2.04	174.31	0.30	98.60	1.10	230.58	0.25	0.34	0.30	1.00
59			(1,3)	1.16	96.36	2.48	160.90	1.16	97.36	1.48	178.08	0.70	97.36	1.94	171.83	0.48	98.48	1.04	223.81	0.36	0.12	0.47	1.00
60			(3,5)	1.32	96.10	2.58	148.06	1.14	97.42	1.44	173.56	0.50	97.54	1.96	164.89	0.60	98.46	0.94	206.05	0.32	0.12	0.59	1.00
61	(100,100)	(20,40)	(1,1)	2.40	95.50	2.10	122.51	1.98	96.50	1.52	131.58	1.68	96.78	1.54	128.30	0.70	98.56	0.74	165.11	-0.07	-0.13	-0.04	0.25
62			(1,3)	2.28	95.80	1.92	116.67	2.30	96.74	0.96	129.62	1.14	97.30	1.56	125.17	0.86	98.54	0.60	156.54	-0.09	-0.41	0.16	1.00
63			(3,5)	2.30	95.52	2.18	107.65	2.20	96.92	0.88	128.12	0.48	98.06	1.46	122.77	0.86	98.46	0.68	143.65	-0.03	-0.43	0.51	1.00
64		(40,40)	(1,1)	2.20	95.52	2.28	133.62	1.78	96.48	1.74	143.42	1.64	96.46	1.90	139.78	0.72	98.46	0.82	180.55	0.02	-0.01	0.07	1.00
65			(1,3)	1.88	95.86	2.26	128.21	2.00	96.68	1.32	141.60	0.94	97.30	1.76	136.82	0.68	98.54	0.78	172.85	0.09	-0.20	0.30	1.00
66			(3,5)	1.58	95.90	2.52	118.25	1.56	97.06	1.38	139.27	0.40	97.62	1.98	133.33	0.44	98.66	0.90	158.17	0.23	-0.06	0.66	1.00
67	(100,200)	(20,40)	(1,1)	1.58	95.16	3.26	101.46	1.34	96.36	2.30	108.80	1.24	96.34	2.42	106.57	0.64	98.36	1.00	134.21	0.35	0.26	0.32	-0.20

(continued on next page)

Table 2 (continued)

Scenarios	(n_1, n_2)	$(\rho_1, \rho_2)\%$	(γ_1, γ_2)	B				F				M				N				Relative bias			
				LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	LEP	CP	UEP	EL	B	F	M	N
68			(1,3)	1.46	95.54	3.00	98.30	1.70	96.74	1.56	108.44	0.88	96.68	2.44	105.68	0.40	98.50	1.10	130.20	0.35	-0.04	0.47	1.00
69			(3,5)	1.68	95.32	3.00	90.42	1.54	97.32	1.14	107.71	0.36	97.30	2.34	104.56	0.56	98.46	0.98	119.13	0.28	-0.15	0.73	1.00
70		(40,40)	(1,1)	1.40	95.62	2.98	114.31	1.04	96.50	2.46	122.60	1.02	96.54	2.44	119.66	0.38	98.56	1.06	152.21	0.36	0.41	0.41	1.00
71			(1,3)	1.28	95.74	2.98	111.45	1.54	96.78	1.68	122.19	0.68	96.70	2.62	118.93	0.36	98.48	1.16	148.24	0.40	0.04	0.59	1.00
72			(3,5)	1.58	95.00	3.42	102.93	1.24	97.16	1.60	121.02	0.48	96.70	2.82	116.90	0.52	98.26	1.22	135.95	0.37	0.13	0.71	1.00

Notes.

B, the Bayesian CI; F, the fiducial GPQ CI; M, the MOVER-fiducial GPQ CI; N, the NA CI.

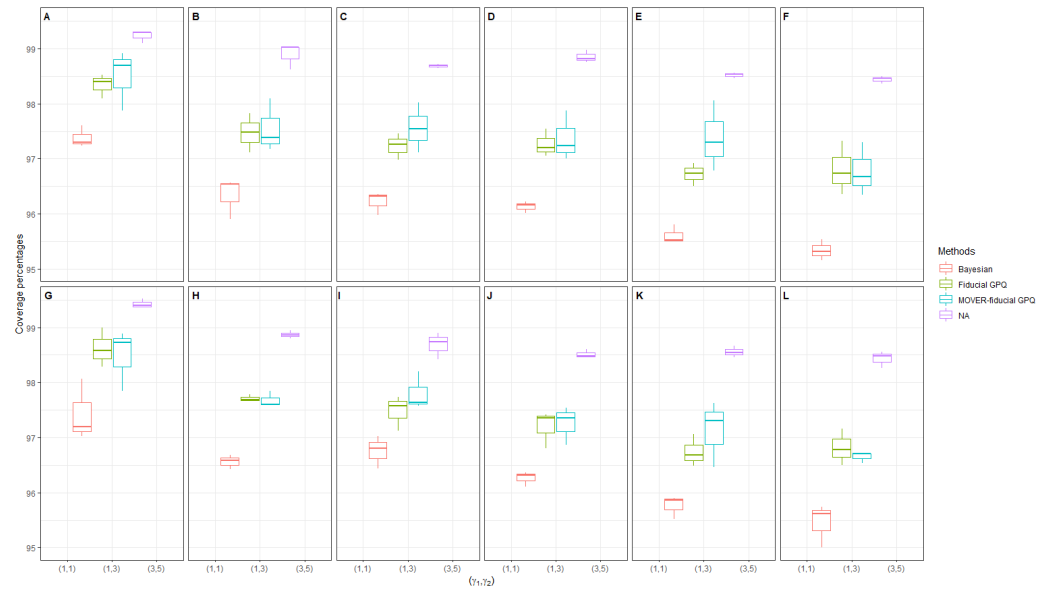


Figure 4 Simulation 2: Box plots of the coverage percentages for Scenarios 37-72 based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 3$ in the following cases [sample sizes, zero proportions]: (A) [(25,25),(20,40)%], (B) [(25,50),(20,40)%], (C) [(50,50),(20,40)%], (D) [(50,100),(20,40)%], (E) [(100,100),(20,40)%], (F) [(100,200),(20,40)%], (G) [(25,25),(40,40)%], (H) [(25,50),(40,40)%], (I) [(50,50),(40,40)%], (J) [(50,100),(40,40)%], (K) [(100,100),(40,40)%], (L) [(100,200),(40,40)%].

Full-size DOI: 10.7717/peerj.14194/fig-4

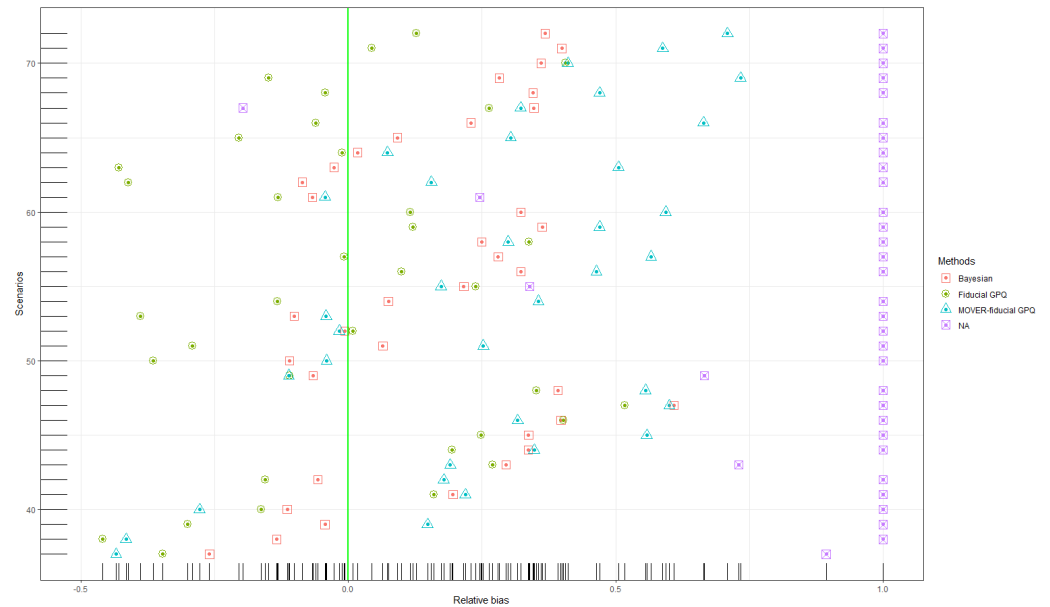


Figure 5 Simulation 2: Scatter plots of the relative bias for Scenarios 37-72 based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 3$.

Full-size DOI: 10.7717/peerj.14194/fig-5

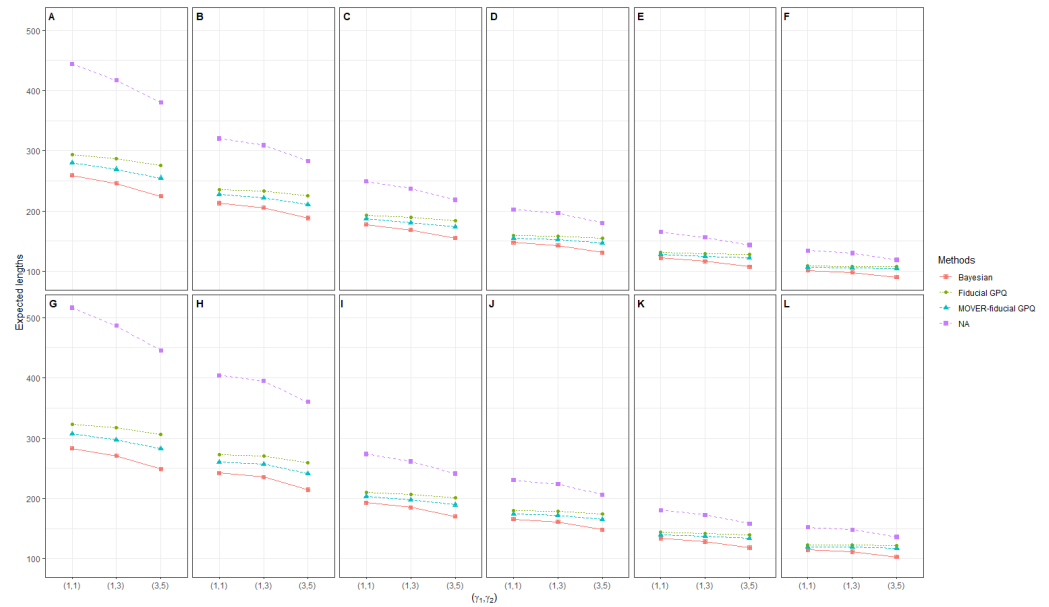


Figure 6 Simulation 2: Line plots of the expected width for 36 scenarios based on the four methods: $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 3$ in the following cases [sample sizes, zero proportions]: (A)[(25,25),(20,40)%],(B)[(25,50),(20,40)%], (C)[(50,50),(20,40)%],(D)[(50,100),(20,40)%], (E)[(100,100),(20,40)%],(F)[(100,200),(20,40)%], (G)[(25,25),(40,40)%],(H)[(25,50),(40,40)%], (I)[(50,50),(40,40)%],(J)[(50,100),(40,40)%], (K)[(100,100),(40,40)%],(L)[(100,200),(40,40)%].

Full-size [DOI: 10.7717/peerj.14194/fig-6](https://doi.org/10.7717/peerj.14194/fig-6)

NA, and Bayesian methods for the ratio of the medians of TPLN distributions containing zero values are reported in Table 6.

It can be interpreted that there is no difference between the wind speeds in Phayao and Phitsanulok. The majority of the population in both areas are agriculturists, and so agricultural burning is often carried out in preparation for planting and after harvesting. Furthermore, the empirical example results are in agreement with the Monte Carlo simulation results in the previous section; the EL of the MOVER-fiducial GPQ CI was the smallest with a suitable CP for a small variance and a large sample size. Overall, the Bayesian-based method is the most suitable for formulating CIs for the ratio of the medians of TPLN distributions containing zero values when taking the checking criteria results from scenarios 1–12 and 37–72 into account.

DISCUSSION

We applied fiducial, NA, and Bayesian-based method to formulate CIs for the ratio of the medians of TPLN distributions containing zero values. From the results of the simulation study, the fiducial methods based on fiducial GPQ and MOVER-fiducial GPQ always provided CPs greater than the nominal 95% confidence level because the fiducial GPQ of $(\gamma_i, \rho_i, \mu_{X_i})$ have strong points (Hannig, Iyer & Patterson, 2006), as revealed by the conditions for FGQ2 in Definition 1. However, the MOVER-fiducial GPQ method produced shorter interval than the fiducial GPQ and worked well for a small variance and a large sample size. The NA method provided CPs greater than the nominal 95% confidence

Table 3 Data of hourly wind speed (km/3 h) in Phisanulok and Phayao provinces, northern Thailand during January 1–15, 2021.

Phisanulok				Phayao			
19.9	0	4.3	13.7	9.4	0	26.7	15
20.6	0	21.6	21.9	16.1	2.2	20.9	13.3
26	11.6	12.6	19.1	17.3	7.5	19.1	9.7
9.4	22.7	18.4	21.2	8.6	15.8	20.5	14
9.8	16.5	16.6	20.1	2.9	14.5	21.7	11.9
12.6	6.1	23.3	20.2	1.8	4.3	29.9	11.1
11.6	0	31	9	0	2.9	33.1	14.3
0	6.8	6.1	22	1.4	4.3	23	12.3
20.9	3.6	1.8	20.9	9.7	2.2	22.6	17.2
14.8	1.5	5	12.6	23.4	0	29.9	6.1
27.3	9	1.8	0	19.8	5.4	24.1	4.7
13	10.4	0	0	3.9	12.6	16.2	3.7
0	18.3	14.4	4.3	1.1	15.4	32	6.1
5.4	8.3	14.5	6.9	0	6.2	29.4	4.6
3.2	0	24.4	16.6	0	3.2	14	11.1
8.7	0.4	6.9	15.9	1.8	2.9	25.2	23.4
8.7	0	2.5	11.6	4.3	0.4	30.2	21.2
13	0	4.3	6.9	22.3	0.4	23.4	4.7
25.9	8.7	2.2	0	21.2	10.8	31.6	1.8
8	25.9	2.5	0	7.6	12.2	29.5	4.7
5	21.7	15.8	0	0.4	21.6	20.1	1.4
3.6	8.3	15.2	2.1	1.1	14.8	30.6	0
5.7	4.7	14.4	4.3	0	13	18.4	7.2
6.2	4.7	7.2	14.4	0.7	17.6	21.2	17.3
7.6	2.2	15.8	19.5	7.2	16.9	14.1	16.6
13	10.1	26.3	5	15.5	2.1	14.8	3.9
25.6	14	32	0	16.2	8.2	19.1	3.2
15.1	20.5	32.7	4.7	5	21.9	25.3	1.8
2.2	17.6	32.4	0	0	21.3	24.5	0.7
3.6	5.8	20.2	4	0	23.5	19.5	0

Notes.

Source: Thai Meteorological Department Automatic Weather System.

URL: http://www.aws-observation.tmd.go.th/web/reports/weather_minute.asp.**Table 4** AIC results for the positive wind speed data (km/3 h) in Phisanulok and Phayao provinces, northern Thailand during January 1–15, 2021.

Provinces	Cauchy	Chi-square	Exponential	Lognormal	TPLN	Logistic	Normal	T
Phitsanulok	787.8140	818.3827	739.9249	733.9261	725.9633	742.5044	736.4505	736.7834
Phayao	865.3148	1008.1470	793.3140	816.9264	800.6639	812.1632	803.3119	803.8294

Notes.

Bold font denotes the best-fitting model for the data.

Table 5 Basic Statistics for the wind speed data.

Basic statistics	Provinces	
	Phitsanulok	Phayao
Groups (i)	1	2
sample sizes (n_i)	120	120
Mean ($\hat{\mu}_i$)	2.1581	2.0196
Variances ($\hat{\sigma}_{x_i}^2$)	1.4911	2.3389
Zero proportions ($\hat{\rho}_i$: %)	13.33	8.33
Threshold ($\hat{\gamma}_i$)	0.399	0.389
Median wind speed ($\hat{\eta}_i$)	8.0248	6.8204
The ratio of the medians ($\hat{\omega} = \hat{\eta}_1 / \hat{\eta}_2$)	$\hat{\omega} = 1.1765$	

Table 6 The 95% CIs for the ratio of the median wind speed in Phisanulok and Phayao provinces.

Methods	95% CIs	Lengths
Fiducial GPQ	(0.7825, 1.7937)	1.0112
MOVER-fiducial GPQ	(0.7823, 1.7203)	0.9379
NA	(0.6922, 1.9997)	1.3074
Bayesian	(0.8047, 1.8238)	1.0190

level, thereby making its ELs longer than the other methods, which could have been caused by the variance in the estimated median (McKean & Schrader, 1984).

Meanwhile, the Bayesian method provided suitable CPs with the shortest interval length, except for a small variance and a large sample size for which the MOVER-fiducial GPQ method performed the best. This could be because of using the uniform prior to gain information about the parameters from the data to obtain their posterior density. As such, the constructed CIs for the ratio of the medians of TPLN distributions containing zero values based on the Bayesian method with a uniform prior performed well.

CONCLUSIONS

CIs for the ratio of the medians of TPLN distributions containing zero values were formulated by using fiducial-, NA-, and Bayesian-based methods. Since a theoretical comparison was not possible, a Monte Carlo simulation and empirical application with two real datasets of wind speed observations were used to evaluate their performances in terms of their CPs and ELs. The results of the simulation study led us to recommend the Bayesian method for constructing the CIs for the ratio of the medians of TPLN distributions containing zero values because it attained CPs close to the nominal 95% confidence level and the shortest EL in most cases, except for a small variance and a large sample size for which the MOVER-fiducial GPQ method should be used.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work (Grant No. RGNS 64–196) was supported by the Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation (OPS MHESI), Thailand Science Research and Innovation (TSRI) and Uttaradit Rajabhat University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation (OPS MHESI): RGNS 64–196.

Thailand Science Research and Innovation (TSRI) and Uttaradit Rajabhat University.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Patcharee Maneerat conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Pisit Nakjai conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Sa-Aat Niwitpong conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.14194#supplemental-information>.

REFERENCES

- Aitchison J, Brown JA. 1963.** *The lognormal distribution: with special reference to its uses in economics*. Cambridge: Cambridge University Press.
- Basak P, Basak I, Balakrishnan N. 2009.** Estimation for the three-parameter lognormal distribution based on progressively censored data. *Computational Statistics & Data Analysis* **53**(10):3580–3592 DOI [10.1016/j.csda.2009.03.015](https://doi.org/10.1016/j.csda.2009.03.015).
- Burges SJ, Lettenmaier DP, Bates CL. 1975.** Properties of the three-parameter log normal probability distribution. *Water Resources Research* **11**(2):229–235 DOI [10.1029/WR011i002p00229](https://doi.org/10.1029/WR011i002p00229).

- Casella G, Berger RL. 2002.** *Statistical inference*. Second Edition. Belmont: Thomson Learning, Inc.
- Chankham W, Niwitpong S-A, Niwitpong S. 2022.** Measurement of dispersion of PM_{2.5} in Thailand using confidence intervals for the coefficient of variation of an inverse Gaussian distribution. *PeerJ* **10**:e12988 DOI [10.7717/peerj.12988](https://doi.org/10.7717/peerj.12988).
- Charbeneau RJ. 1978.** Comparison of the two- and three-parameter log normal distributions used in streamflow synthesis. *Water Resources Research* **14**(1):149–150 DOI [10.1029/WR014i001p00149](https://doi.org/10.1029/WR014i001p00149).
- Chen Z, Miao F. 2012.** Interval and point estimators for the location parameter of the three-parameter lognormal distribution. *International Journal of Quality, Statistics, and Reliability* **2012**:1–6.
- Cohen AC, Whitten BJ. 1980.** Estimation in the three-parameter lognormal distribution. *Journal of the American Statistical Association* **75**(370):399–404 DOI [10.1080/01621459.1980.10477484](https://doi.org/10.1080/01621459.1980.10477484).
- Cohen AC, Whitten BJ, Ding Y. 1985.** Modified moment estimation for the three-parameter lognormal distribution. *Journal of Quality Technology* **17**(2):92–99 DOI [10.1080/00224065.1985.11978942](https://doi.org/10.1080/00224065.1985.11978942).
- Donner A, Zou GY. 2012.** Closed-form confidence intervals for functions of the normal mean and standard deviation. *Statistical Methods in Medical Research* **21**(4):347–359 DOI [10.1177/0962280210383082](https://doi.org/10.1177/0962280210383082).
- Fisher RA. 1935.** The fiducial argument in statistical inference. *Annals of Eugenics* **6**(4):391–398 DOI [10.1111/j.1469-1809.1935.tb02120.x](https://doi.org/10.1111/j.1469-1809.1935.tb02120.x).
- Griffiths DA. 1980.** Interval estimation for the three-parameter lognormal distribution via the likelihood function. *Applied Statistics* **29**(1):58–68 DOI [10.2307/2346411](https://doi.org/10.2307/2346411).
- Group BP. 2021.** RISC by MQDC and University of Phayao collaborate in PM_{2.5} research in northern Thailand. Bangkok Post.
- Hannig J, Iyer H, Patterson P. 2006.** Fiducial generalized confidence intervals. *Journal of the American Statistical Association* **101**(473):254–269 DOI [10.1198/016214505000000736](https://doi.org/10.1198/016214505000000736).
- Harvey J, Van der Merwe AJ. 2012.** Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. *Journal of Statistical Planning and Inference* **142**(6):1294–1309 DOI [10.1016/j.jspi.2011.12.006](https://doi.org/10.1016/j.jspi.2011.12.006).
- Hasan MS, Krishnamoorthy K. 2017.** Improved confidence intervals for the ratio of coefficients of variation of two lognormal distributions. *Journal of Statistical Theory and Applications* **16**(3):345–353 DOI [10.2991/jsta.2017.16.3.6](https://doi.org/10.2991/jsta.2017.16.3.6).
- Hettmansperger TP, Sheather SJ. 1986.** Confidence intervals based on interpolated order statistics. *Statistics & Probability Letters* **4**(2):75–79 DOI [10.1016/0167-7152\(86\)90021-0](https://doi.org/10.1016/0167-7152(86)90021-0).
- IQAir. 2021.** New data exposes Thailand’s 2021 burning season. Available at <https://www.iqair.com/blog/air-quality/thailand-2021-burning-season>.
- IQAir. 2022.** Air quality in Phitsanulok. Available at <https://www.iqair.com/th-en/thailand/phitsanulok>.

- Kharrati-Kopaei M, Malekzadeh A, Sadooghi-Alvandi M. 2013.** Simultaneous fiducial generalized confidence intervals for the successive differences of exponential location parameters under heteroscedasticity. *Statistics & Probability Letters* **83(6)**:1547–1552 DOI [10.1016/j.spl.2013.03.001](https://doi.org/10.1016/j.spl.2013.03.001).
- Krishnamoorthy K, Mathew T. 2003.** Inferences on the means of lognormal distributions using generalized p -values and generalized confidence intervals. *Journal of Statistical Planning and Inference* **115(1)**:103–121 DOI [10.1016/S0378-3758\(02\)00153-2](https://doi.org/10.1016/S0378-3758(02)00153-2).
- Li X, Zhou X, Tian L. 2013.** Interval estimation for the mean of lognormal data with excess zeros. *Statistics & Probability Letters* **83(11)**:2447–2453 DOI [10.1016/j.spl.2013.07.004](https://doi.org/10.1016/j.spl.2013.07.004).
- Liu Z, Shen L, Yan C, Du J, Li Y, Zhao H. 2020.** Analysis of the influence of precipitation and wind on PM2.5 and PM10 in the atmosphere. *Advances in Meteorology* **2020**:e5039613.
- Maneerat P, Nakjai P, Niwitpong S-A. 2022a.** Bayesian interval estimations for the mean of delta-three parameter lognormal distribution with application to heavy rainfall data. *PLOS ONE* **17(4)**:e0266455 DOI [10.1371/journal.pone.0266455](https://doi.org/10.1371/journal.pone.0266455).
- Maneerat P, Niwitpong S-A. 2020.** Comparing medical care costs using bayesian credible intervals for the ratio of means of delta-lognormal distributions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **28((Supp 01))**:51–68 DOI [10.1142/S021848852040005X](https://doi.org/10.1142/S021848852040005X).
- Maneerat P, Niwitpong S-A, Nakjai P. 2022b.** Estimating wind speed by using confidence intervals for the median in a three-parameter lognormal model. In: Honda K, Entani T, Ubukata S, Huynh V-N, Inuiguchi M, eds. *Integrated uncertainty in knowledge modelling and decision making. Lecture Notes in Computer Science*, Cham: Springer International Publishing., 329–341.
- Maneerat P, Niwitpong S, Niwitpong S-A. 2021.** Bayesian confidence intervals for variance of delta-lognormal distribution with an application to rainfall dispersion. *Statistics and its Interface* **14**:229–241 DOI [10.4310/20-SII630](https://doi.org/10.4310/20-SII630).
- Maneerat P, Niwitpong S-A, Niwitpong S. 2020.** A Bayesian approach to construct confidence intervals for comparing the rainfall dispersion in Thailand. *PeerJ* **8**:e8502 DOI [10.7717/peerj.8502](https://doi.org/10.7717/peerj.8502).
- McKean JW, Schrader RM. 1984.** A comparison of methods for studentizing the sample median. *Communications in Statistics—Simulation and Computation* **13(6)**:751–773 DOI [10.1080/03610918408812413](https://doi.org/10.1080/03610918408812413).
- Pang W-K, Leung P-K, Huang W-K, Liu W. 2005.** On interval estimation of the coefficient of variation for the three-parameter Weibull, lognormal and gamma distribution: a simulation-based approach. *European Journal of Operational Research* **164(2)**:367–377 DOI [10.1016/j.ejor.2003.04.005](https://doi.org/10.1016/j.ejor.2003.04.005).
- Royston P. 1992.** Estimation, reference ranges and goodness of fit for the three-parameter log-normal distribution. *Statistics in Medicine* **11(7)**:897–912 DOI [10.1002/sim.4780110707](https://doi.org/10.1002/sim.4780110707).

- Ruder S. 2016.** An overview of gradient descent optimization algorithms. ArXiv preprint. DOI 10.48550/arXiv.1609.04747.
- Singh G. 2020.** From SGD to Adam. Available at <https://medium.com/mdr-inc/from-sgd-to-adam-c9fce513c4bb>.
- Singh V, Cruise J, Ma M. 1990.** A comparative evaluation of the estimators of the three-parameter lognormal distribution by Monte Carlo simulation. *Computational Statistics & Data Analysis* 10(1):71–85 DOI 10.1016/0167-9473(90)90104-P.
- Singh VP, Rajagopal AK. 1986.** A new method of parameter estimation for hydrologic frequency analysis. *Hydrological Science and Technology* 2(3):33–40.
- Singh VP, Singh K. 1987.** Parameter estimation for TPLN distribution for flood frequency analysis. *Journal of the American Water Resources Association* 23(6):1185–1189 DOI 10.1111/j.1752-1688.1987.tb00871.x.
- Tanraksa P, Kendall D. 2020.** Air quality in North still critical. Bangkok Post.
- Teerasuphaset T, Culp J. 2020.** *Stay safe in the PM 2.5*. Bangkok: Chulalongkorn University.
- Thai Meteorological Department Automatic Weather System. 2022.** Reports of gathering data by time slot. Available at http://www.aws-observation.tmd.go.th/web/reports/weather_minute.asp.
- Weerahandi S. 1993.** Generalized confidence intervals. *Journal of the American Statistical Association* 88(423):899–905 DOI 10.1080/01621459.1993.10476355.
- Wipatayotin A, Tangprasert P. 2021.** PM2.5 dust spikes in North, Northeast. Bangkok Post. Available at <https://www.bangkokpost.com/thailand/general/2057587/pm2-5-dust-spikes-in-north-northeast>.
- Yosboonruang N, Niwitpong S-A, Niwitpong S. 2022.** Bayesian computation for the common coefficient of variation of delta-lognormal distributions with application to common rainfall dispersion in Thailand. *PeerJ* 10:e12858 DOI 10.7717/peerj.12858.
- Zhang Q, Xu J, Zhao J, Liang H, Li X. 2021.** Simultaneous confidence intervals for ratios of means of zero-inflated log-normal populations. *Journal of Statistical Computation and Simulation* 92(6):1113–1132 DOI 10.1080/00949655.2021.1986508.