Confidence intervals for the common coefficient of variation of rainfall in Thailand

Warisa Thangjai [1] , **Sa-Aat Niwitpong** [Corresp., 2] , **Suparat Niwitpong** [3]

1 Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand

2 Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangok, Bangkok, Bangkok, Thailand

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

Corresponding Author: Sa-Aat Niwitpong
Email address: sa-aat.n@sci.kmutnb.ac.th

The log-normal distribution is often used to analyze environmental data like daily rainfall amounts. The rainfall is of interest in Thailand because high variable climates can lead to periodic water stress and scarcity. The mean, standard deviation or coefficient of variation of the rainfall in the area is usually estimated. The climate moisture index is the ratio of plant water demand to precipitation. The climate moisture index should use the coefficient of variation instead of the standard deviation for comparison between areas with widely different means. The larger coefficient of variation indicates greater dispersion, whereas the lower coefficient of variation indicates the lower risk. The common coefficient of variation, is the weighted coefficients of variation based on k areas, presents the average daily rainfall. Therefore, the common coefficient of variation is used to describe overall water problems of k areas. In this paper, we propose four novel approaches for the confidence interval estimation of the common coefficient of variation of log-normal distributions based on the fiducial generalized confidence interval (FGCI), method of variance estimates recovery (MOVER), computational, and Bayesian approaches. A Monte Carlo simulation was used to evaluate the coverage probabilities and average lengths of the confidence intervals. In terms of coverage probability, the results show that the FGCI approach provided the best confidence interval estimates for most cases except for when the sample case was equal to six populations ($k =$ 6) and the sample sizes were small ($n_{i} <$ 50), for which the MOVER confidence interval estimates were the best. The efficacies of the proposed approaches are illustrated with example using real-life daily rainfall datasets from regions of Thailand.

Confidence Intervals for the Common Coefficient of Variation of Rainfall in Thailand

Warisa Thangjai[1], Sa-Aat Niwitpong[2], and Suparat Niwitpong[2]

[1]Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok, 10240, Thailand
[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand

Corresponding author:

Sa-Aat Niwitpong[a]

Email address: sa-aat.n@sci.kmutnb.ac.th

ABSTRACT

The log-normal distribution is often used to analyze environmental data like daily rainfall amounts. The rainfall is of interest in Thailand because high variable climates can lead to periodic water stress and scarcity. The mean, standard deviation or coefficient of variation of the rainfall in the area is usually estimated. The climate moisture index is the ratio of plant water demand to precipitation. The climate moisture index should use the coefficient of variation instead of the standard deviation for comparison between areas with widely different means. The larger coefficient of variation indicates greater dispersion, whereas the lower coefficient of variation indicates the lower risk. The common coefficient of variation, is the weighted coefficients of variation based on k areas, presents the average daily rainfall. Therefore, the common coefficient of variation is used to describe overall water problems of k areas. In this paper, we propose four novel approaches for the confidence interval estimation of the common coefficient of variation of log-normal distributions based on the fiducial generalized confidence interval (FGCI), method of variance estimates recovery (MOVER), computational, and Bayesian approaches. A Monte Carlo simulation was used to evaluate the coverage probabilities and average lengths of the confidence intervals. In terms of coverage probability, the results show that the FGCI approach provided the best confidence interval estimates for most cases except for when the sample case was equal to six populations ($k = 6$) and the sample sizes were small ($n_i < 50$), for which the MOVER confidence interval estimates were the best. The efficacies of the proposed approaches are illustrated with example using real-life daily rainfall datasets from regions of Thailand.

INTRODUCTION

Droughts and floods are regular natural disasters in Thailand. Droughts occur when the hot season begins after a year with unusually little rainfall. Moreover, floods happen nearly every year during the monsoon season. The monsoon seasons in the country are distinct by region. Thailand is divided into six geographical regions such as the north, the northeast, the west, the central, the east, and the south. Various regions are prone to seasonal flash-flooding. The floods often occur in the north, the northern east, and the south. Since rainfall varies greatly depending on region and season. Therefore, the common coefficient of variation is used to represent the rainfall dispersion in different regions.

The log-normal distribution, widely used to describe the distribution of right-skewed data, has been used to model various real-life applications (Jafari and Abdollahnezhad, 2015). For instance, in climate sciences and hydrology, the rainfall measurements are right-skewed (Thangjai et al., 2020a). The coefficient of variation of the log-normal distribution depends on the variance (σ^2) only (Thangjai et al, 2016) whereas the coefficient of variation of normal distribution depends on the mean (μ) and σ^2. Although the normal distribution is more well-known than the log-normal distribution in the natural and social sciences, the latter has been used in many applications. Examples of quantities which have

approximate log-normal distributions include the particulate matter and rainfall frequency. Rumburg et al. (2001) studied the statistical distributions of daily particulate matter data from Spokane, Washington from January 1995 to December 1997. They found that the PM2.5 data is best fit by a three parameter log-normal distribution. Thangjai et al. (2020a) constructed the simultaneous confidence intervals for all differences of coefficients of variation of daily rainfall data on 17 July 2018 in different regions of Thailand. The daily rainfall data is log-normal distribution.

In statistic, the information in a sample $X = (X_1, X_2, \ldots, X_n)$ is used to make inferences about an unknown parameter θ. The inference methods are hypothesis testing, point estimation, and confidence interval estimation. The statistical hypothesis makes a statement about a population parameter. The hypothesis testing uses the sample from the population for deciding. Two complementary hypotheses in hypothesis testing are the null hypothesis and the alternative hypothesis. The point estimation uses the sample data to evaluate a single value. The point estimation is a guess of the single value as the value of the parameter. The value is called the point estimator. The point estimate is not absolutely accurate because the estimate is based on only the single random sample. For a contrasting point estimation method, the confidence interval estimation uses the sample data to calculate an interval of probable values. The confidence interval is called the confidence interval estimator. The confidence interval estimation is used rather than the point estimation because the confidence interval estimation has some guarantee of capturing the parameter. The goal of this paper is to examine the confidence interval for parameter of log-normal distribution. Confidence intervals associated with various functions of the log-normal distribution parameters have been reported by Land (1988), Zhou and Gao (1997), Zhou (1998), Joulious and Debarnot (2000), Taylor et al. (2002), Krishnamoorthy and Mathew (2003), Gupta and Li (2005), Hannig et al. (2006b), and Shen et al. (2006). In these studies, confidence intervals were considered on linear functions of the mean and variance of the log-normal distribution. These results were later extended to stochastic processes such as homogeneous log-normal (Gutiérrez et al., 2007) and non-homogeneous log-normal (Gutiérrez et al., 2003).

The coefficient of variation is defined as the standard deviation divided by the mean (Kelley, 2007). It can be used to compare several populations that have different measurement units and is widely used to measure the precision and repeatability of data in many fields. In hematology and serology, the coefficient of variation has been used for the measurement of blood samples taken from different laboratories (Tsim et al., 1991) and as a measure of precision within and between laboratories (Tian, 2005). In finance, a test of the equality of the coefficients of variation for two stocks has been used to measure risk. In medicine, the coefficient of variation has been used to compare the variability in the ratio of total/high density lipoprotein cholesterol with the variability in vessel diameter change according to diet.

In climate sciences and hydrology, the coefficient of variation has been used to describe the rainfall and can be used to compare the rainfall variability in two or more different areas (Thangjai et al., 2020a). If the difference between the rainfall in a single area and the average rainfall over several areas is high, then the rainfall is high. In statistical analysis, combine the results of several independent studies is used in climate sciences and clinical trial. If it is assumed that the samples are collected from independent log-normal populations with a common coefficient of variation but possibly with different variances, then the confidence interval for the common coefficient of variation of several log-normal populations becomes the parameter of interest. Several researchers have focused on confidence interval estimation for the coefficient of variation of a log-normal distribution. For example, Niwitpong (2013) presented confidence intervals for the coefficient of variation of a log-normal distribution with a restricted parameter space. Ng (2014) proposed an approach to make inference on the common coefficient of variation of log-normal populations. Simultaneous confidence intervals for the differences in the coefficients of variation of log-normal distributions were proposed by Thangjai et al. (2016) and Thangjai et al. (2020a). Moreover, Nam and Kwon (2017) and Hasan and Krishnamoorthy (2017) studied the confidence intervals for the ratio of coefficients of variation of two log-normal distributions.

Common problems in applied statistics are confidence interval estimation for the coefficient of variation and testing the equality of two or more coefficients of variation. Miller and Karson (1977) proposed a test for the equality of coefficients of variation in two normal populations. Gupta and Ma (1996) presented testing the equality of the coefficients of variation in k normal populations. Fung and Tsang (1998) compared parametric and nonparametric tests and Gokpinar and Gokpinar (2015) proposed a computational approach to test the equality of the coefficients of variation of k normal populations. Sangnawakij (2016) proposed two new confidence intervals for the ratio of the coefficients of variation

of two-parameter exponential distributions. Sangnawakij and Niwitpong (2016) presented confidence interval estimation for the single coefficient of variation and the difference between two coefficients of variation in two-parameter exponential distributions.

Under many circumstances, confidence interval estimation or hypothesis testing for the common coefficient of variation based on several independent samples is of interest (Tian, 2005). Krishnamoorthy and Lu (2003) investigated procedures for confidence interval estimation and hypothesis testing of the common mean of several normal populations, while the problem of making inference from common populations with a common coefficient of variation of normal distributions was dealt with by Tian (2005). Tian and Wu (2007) proposed confidence interval estimation and hypothesis testing of the common mean of several log-normal populations using the generalized variable concept. Similarly, procedures for hypothesis testing and confidence interval estimation for the common mean of several inverse Gaussian populations were presented by Ye et al. (2010). Moreover, Ng (2014) constructed a confidence interval for the common coefficient of variation from several independent log-normal samples based on the GCI approach, although there was no comparison with other approaches.

The concepts of the generalized pivotal quantity (GPQ) and the GCI first proposed by Weerahandi (1993) have been applied to solve many statistical problems. For example, Krishnamoorthy and Lu (2003) presented the generalized variable and GCI approaches for inference on the common mean of several normal populations. Tian (2005) studied inferences on the common coefficient of variation of several normal populations. Moreover, Tian and Wu (2007) proposed the generalized variable approach and the GCI approach for inferences on the common mean of several log-normal populations. Thangjai et al. (2020b) developed confidence intervals for the common coefficient of variation of several normal populations using the GCI and adjusted GCI approaches. Hannig et al. (2006a) suggested fiducial GPQ (FGPQ) as a subclass of GPQ, with the FGCI being constructed using the FGPQ under fairly general conditions. Furthermore, FGCIs have been constructed to solve many practical problems (Hannig et al., 2006b; Chang and Huang, 2009; Kharrati-Kopaei et al. 2013; Thangjai et al., 2016). Although the FGCI approach is based on simulated data, one advantage is that it can be used to construct the confidence interval for complex parameters.

The method of variance estimates recovery (MOVER) approach was introduced by Zou and Donner (2008) and Zou et al. (2009). Several researchers have successfully used the MOVER approach to construct confidence intervals (Donner and Zou, 2012; Suwan and Niwitpong, 2013; Niwitpong and Wongkhao, 2016). The MOVER approach has the advantage of being easy to compute using the exact formula. However, the disadvantage of this approach is that one can construct it with or without the initial confidence interval for a single parameter of interest.

The computational approach proposed by Pal et al. (2007) has been used by many researchers to test the equality of several populations (e.g. Jafari and Abdollahnezhad, 2015; Jafari and Kazemi, 2017; Gokpinar and Gokpinar, 2017). As an advantage, this approach does not require explicit knowledge of the sampling distribution of the test statistic. However, it is based on simulation and numerical computations using the maximum likelihood estimate only. The Bayesian approach uses Bayes' theorem to compute and update probabilities after obtaining new data. Although it can be applied to estimate the confidence intervals for complex parameters, the disadvantages of applying it are that it requires prior information and is based on simulation.

Ng (2014) constructed the GCI of a log-normal distribution which used asymptotic variance and provided coverage probabilities close to the nominal confidence level of 0.95. The confidence interval constructed by transforming the log-normal coefficient of variation to the normal coefficient of variation. However, its use was only considered for homogeneous populations. Therefore, in this study, we extend the research of Ng (2014) to develop four novel approaches for confidence interval estimation for the common coefficient of variation of several log-normal populations based on the FGCI, MOVER, computational approach, and Bayesian approaches. Unlike Ng (2014), we compute the confidence intervals for the common coefficient of variation directly using the log-normal coefficient of variation that depends on σ^2 only. Moreover, there is no previous literature on applying their methodology to PM2.5 concentration measurements. Therefore, to fill the gap, the novel approaches for the confidence interval estimation of the common coefficient of variation of log-normal distributions were proposed with considering the log-normality of PM2.5 concentration measurements.

METHODS

Let us consider Y the log-normal distribution with parameters μ_Y and σ_Y^2. It is well known that $X = \log(Y)$ follows a normal distribution with mean μ and variance σ^2, whereas the mean and variance of Y are given by

$$\mu_Y = \exp\left(\mu + \left(\sigma^2/2\right)\right) \tag{1}$$

and

$$\sigma_Y^2 = \left(\exp\left(\sigma^2\right) - 1\right) \cdot \left(\exp\left(2\mu + \sigma^2\right)\right), \tag{2}$$

respectively. From Equation (1) and Equation (2), the coefficient of variation of Y is given by

$$\theta = \sqrt{\exp\left(\sigma^2\right) - 1}. \tag{3}$$

From Equation (3), it is seen that the coefficient of variation of the log-normal distribution depends on parameter σ^2 only, whereas the coefficient of variation of the normal distribution depends on μ and σ^2. And the next result provides a useful approximation for the variance of an estimator of θ.

Let $\hat{\theta} = \sqrt{\exp\left(S^2\right) - 1}$ be an estimator of θ. Following the Niwitpong (2013) and Hasan and Krishnamoorthy (2017), the variance of $\hat{\theta}$ is

$$Var\left(\hat{\theta}\right) \approx \frac{\sigma^4 \exp\left(2\sigma^2\right)}{2\left(n-1\right) \cdot \left(\exp\left(\sigma^2\right) - 1\right)}. \tag{4}$$

Suppose that random samples are taken from k log-normal distributions, $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i}) \sim LN(\mu_i, \sigma_i^2)$, where $i = 1, 2, \ldots, k$. Let $\theta_i = \sqrt{\exp\left(\sigma_i^2\right) - 1}$ be the coefficient of variation for $i = 1, 2, \ldots, k$. Let $X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i})$ be a random variable of size n_i from the normal distribution with μ_i and variance σ_i^2. Let $\hat{\mu}_i = \bar{X}_i$ and $\hat{\sigma}_i^2 = S_i^2$ be estimators of μ_i and σ_i^2, respectively, where $\bar{X}_i$ and S_i^2 denote the mean and variance of the log-transformed sample from a log-normal distribution. The mean and variance are given by

$$\bar{X}_i = \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \tag{5}$$

and

$$S_i^2 = \frac{\sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_i\right)^2}{n_i - 1}, \tag{6}$$

where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$.

Let $\bar{x}_i$ and s_i^2 be the observed values of $\bar{X}_i$ and S_i^2, respectively. The maximum likelihood estimator of the coefficient of variation θ_i, is also unbiased estimator, is given by

$$\hat{\theta}_i = \sqrt{\exp\left(S_i^2\right) - 1}, \tag{7}$$

where S_i^2 is defined in Equation (6).

Fiducial generalized confidence interval

The FGCI uses the FGPQs. The FGPQs are a subclass of the GPQs. The FGCI has correct frequentist coverage probability. For $i = 1, 2, \ldots, k$, let $\bar{X}_i$ and S_i^2 be the sample mean and the sample variance for log-transformed data and let $\bar{x}_i$ and s_i^2 be the observed sample mean and the observed sample variance respectively. Let $\bar{X}_i^*$ and S_i^{2*} be independent copies of $\bar{X}_i$ and S_i^2, respectively. Let $\bar{X}_i$ and $\bar{X}_i^*$ be independent and identically distributed with mean μ_i and variance σ_i^2/n_i. It is well known that

$$\bar{X}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{n_i}\right) \quad \text{and} \quad \bar{X}_i^* \sim N\left(\mu_i, \frac{\sigma_i^2}{n_i}\right). \tag{8}$$

Furthermore, let S_i^2 and S_i^{2*} be independent and identically distributed. Then

$$\frac{(n_i-1)\,S_i^2}{\sigma_i^2} \sim \chi_{n_i-1}^2 \quad \text{and} \quad \frac{(n_i-1)\,S_i^{2*}}{\sigma_i^2} \sim \chi_{n_i-1}^2, \tag{9}$$

where $\chi_{n_i-1}^2$ is chi-squared distribution with $n_i - 1$ degree of freedom.

According to Hannig et al. (2006a, 2006b), the FGPQs of μ_i and σ_i^2 are defined by

$$R_{\mu_i} = \bar{X}_i - \frac{S_i}{S_i^*}\left(\bar{X}_i^* - \mu_i\right) \tag{10}$$

and

$$R_{\sigma_i^2} = \frac{S_i^2}{S_i^{2*}}\sigma_i^2. \tag{11}$$

Therefore, the FGPQ for θ_i based on the FGPQ for σ_i^2 is given by

$$R_{\theta_i} = \sqrt{\exp\left(R_{\sigma_i^2}\right) - 1}. \tag{12}$$

The FGPQ of θ_i in Equation (12) satisfies two conditions defined in Definition of Hannig et al. (2006a, 2006b). The definition of Hannig et al. (2006a, 2006b) has two conditions such as the distribution of the GPQ is free of all unknown parameters and the observed value of the GPQ is the parameter of interest. From Equation (4), the variance of $\hat{\theta}_i$ is provided by

$$Var\left(\hat{\theta}_i\right) \approx \frac{\sigma_i^4 \exp\left(2\sigma_i^2\right)}{2\left(n_i-1\right)\cdot\left(\exp\left(\sigma_i^2\right)-1\right)}. \tag{13}$$

The FGPQ of $Var\left(\hat{\theta}_i\right)$ is given by

$$R_{Var(\hat{\theta}_i)} = \frac{R_{\sigma_i^2}^2 \exp\left(2R_{\sigma_i^2}\right)}{2\left(n_i-1\right)\cdot\left(\exp\left(R_{\sigma_i^2}\right)-1\right)}. \tag{14}$$

The FGPQ for the common coefficient of variation θ is a weighted average of the FGPQ R_{θ_i} based on k individual sample. Therefore, the FGPQ is given by

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

where R_{θ_i} is defined in Equation (12) and $R_{Var(\hat{\theta}_i)}$ is defined in Equation (14).

The FGPQ in Equation (12) satisfies two conditions of the definition given above. The FGCI is constructed using the quantiles of FGPQ defined in Equation (15). Therefore, the $100\,(1-\alpha)\,\%$ two-sided confidence interval for the common coefficient of variation θ based on the FGCI approach is

$$CI_{FGCI} = [L_{FGCI}, U_{FGCI}] = [R_\theta\left(\alpha/2\right), R_\theta\left(1-\alpha/2\right)], \tag{16}$$

where $R_\theta\left(\alpha/2\right)$ and $R_\theta\left(1-\alpha/2\right)$ denote the $100\,(\alpha/2)$-th and $100\,(1-\alpha/2)$-th percentiles of R_θ, respectively.

The following algorithm is used to construct the FGCI:

Algorithm 1

For a given $\bar{x}_i$ and s_i^2, where $i = 1, 2, \ldots, k$

For $g = 1$ to m, where m is number of generalized computation

Generate X^* and then compute $\bar{x}_i^*$ and s_i^{2*}

Generate $\chi_{n_i-1}^2$ from chi-squared distribution with $n_i - 1$ degrees of freedom

Compute $R_{\sigma_i^2}$ from Equation (11)

Compute R_{θ_i} from Equation (12)

Compute $R_{Var(\hat{\theta}_i)}$ from Equation (14)

Compute R_θ from Equation (15)

End g loop

Compute $R_\theta\left(\alpha/2\right)$ and $R_\theta\left(1-\alpha/2\right)$ from Equation (16)

Method of variance estimates recovery confidence interval

Zou and Donner (2008) and Zou et al. (2009) proposed the MOVER approach to construct the confidence interval for the sum of two parameters. For $i = 1, 2$, let θ_1 and θ_2 be the parameters of interest. Let L and U be the lower limit and upper limit of the confidence interval for $\theta_1 + \theta_2$. Moreover, let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the estimators of θ_1 and θ_2, respectively. The central limit theorem and the assumption of independence between the point estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are used. Therefore, the lower limit L is

$$L = \hat{\theta}_1 + \hat{\theta}_2 - z_{\alpha/2} \sqrt{\widehat{Var}\left(\hat{\theta}_1\right) + \widehat{Var}\left(\hat{\theta}_2\right)}, \tag{17}$$

where $z_{\alpha/2}$ is the $100\left(\alpha/2\right)$-th percentile of the standard normal distribution.

Let l_i and u_i be the lower limit and upper limit of the confidence interval for θ_i, where $i = 1, 2$. The lower limit L must be closer to $l_1 + l_2$ than to $\hat{\theta}_1 + \hat{\theta}_2$. The variance estimate for $\hat{\theta}_i$ at $\theta_i = l_i$ is defined by

$$\widehat{Var}\left(\hat{\theta}_{l_i}\right) = \frac{\left(\hat{\theta}_i - l_i\right)^2}{z_{\alpha/2}^2}. \tag{18}$$

Substituting back into Equation (17) as follows

$$L = \hat{\theta}_1 + \hat{\theta}_2 - \sqrt{\left(\hat{\theta}_1 - l_1\right)^2 + \left(\hat{\theta}_2 - l_2\right)^2}. \tag{19}$$

Similarly, the variance estimate for $\hat{\theta}_i$ at $\theta_i = u_i$ is defined by

$$\widehat{Var}\left(\hat{\theta}_{u_i}\right) = \frac{\left(u_i - \hat{\theta}_i\right)^2}{z_{\alpha/2}^2}. \tag{20}$$

The upper limit U is

$$U = \hat{\theta}_1 + \hat{\theta}_2 + \sqrt{\left(u_1 - \hat{\theta}_1\right)^2 + \left(u_2 - \hat{\theta}_2\right)^2}. \tag{21}$$

Therefore, the variance estimate for $\hat{\theta}_i$ at $\theta_i = l_i$ and $\theta_i = u_i$ is defined by

$$\widehat{Var}\left(\hat{\theta}_i\right) = \frac{1}{2}\left(\frac{\left(\hat{\theta}_i - l_i\right)^2}{z_{\alpha/2}^2} + \frac{\left(u_i - \hat{\theta}_i\right)^2}{z_{\alpha/2}^2}\right), \tag{22}$$

where $i = 1, 2$.

In this paper, the k parameters of interest are $\theta_1, \theta_2, \ldots, \theta_k$. The concepts of Zou and Donner (2008) and Zou et al. (2009) are motivated for constructing the confidence interval for $\theta_1 + \theta_2 + \ldots + \theta_k$ are

$$L = \hat{\theta}_1 + \hat{\theta}_2 + \ldots + \hat{\theta}_k - \sqrt{\left(\hat{\theta}_1 - l_1\right)^2 + \left(\hat{\theta}_2 - l_2\right)^2 + \ldots + \left(\hat{\theta}_k - l_k\right)^2} \tag{23}$$

and

$$U = \hat{\theta}_1 + \hat{\theta}_2 + \ldots + \hat{\theta}_k + \sqrt{\left(u_1 - \hat{\theta}_1\right)^2 + \left(u_2 - \hat{\theta}_2\right)^2 + \ldots + \left(u_k - \hat{\theta}_k\right)^2}, \tag{24}$$

where $(l_1, u_1), (l_2, u_2), \ldots, (l_k, u_k)$ contain the parameter values for $\theta_1, \theta_2, \ldots, \theta_k$, respectively.

According to Graybill and Deal (1959), the common coefficient of variation θ is weighted average of the coefficient of variation $\hat{\theta}_i$ based on k individual samples. Therefore, the common coefficient of variation is defined by

$$\hat{\theta} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{\widehat{Var}\left(\hat{\theta}_i\right)} \bigg/ \sum_{i=1}^{k} \frac{1}{\widehat{Var}\left(\hat{\theta}_i\right)}, \tag{25}$$

where $\hat{\theta}_i = \sqrt{\exp\left(S_i^2\right) - 1}$ and $\widehat{Var}\left(\hat{\theta}_i\right)$ is defined in Equation (22).

Applying Krishnamoorthy and Oral (2017), the lower limit and upper limit of the confidence interval for the common coefficient of variation θ are defined by

$$L_{MOVER} = \hat{\theta} - \sqrt{\sum_{i=1}^{k} \frac{\left(\hat{\theta}_i - l_i\right)^2}{\left(\widehat{Var}\left(\hat{\theta}_{l_i}\right)\right)^2} \Bigg/ \sum_{i=1}^{k} \frac{1}{\left(\widehat{Var}\left(\hat{\theta}_{l_i}\right)\right)^2}} \qquad (26)$$

and

$$U_{MOVER} = \hat{\theta} + \sqrt{\sum_{i=1}^{k} \frac{\left(u_i - \hat{\theta}_i\right)^2}{\left(\widehat{Var}\left(\hat{\theta}_{u_i}\right)\right)^2} \Bigg/ \sum_{i=1}^{k} \frac{1}{\left(\widehat{Var}\left(\hat{\theta}_{u_i}\right)\right)^2}}, \qquad (27)$$

where $\hat{\theta}$ is defined in Equation (25).

According to Niwitpong (2013), for $i = 1, 2, \ldots, k$, the confidence interval for coefficient of variation of log-normal distribution based on the i-th sample is given by

$$[l_i, u_i] = \left[\sqrt{\exp\left(\frac{(n_i-1)S_i^2}{\chi^2_{(n_i-1),(1-\alpha/2)}}\right) - 1}, \sqrt{\exp\left(\frac{(n_i-1)S_i^2}{\chi^2_{(n_i-1),(\alpha/2)}}\right) - 1}\right], \qquad (28)$$

where $\chi^2_{(n_i-1),(1-\alpha/2)}$ and $\chi^2_{(n_i-1),(\alpha/2)}$ denote the $100\left(1-\alpha/2\right)$-th and $100\left(\alpha/2\right)$-th percentiles of the chi-squared distribution with $n_i - 1$ degrees of freedom.

Therefore, the $100\left(1-\alpha\right)\%$ two-sided confidence interval for the common coefficient of variation θ based on MOVER approach is

$$CI_{MOVER} = [L_{MOVER}, U_{MOVER}], \qquad (29)$$

where L_{MOVER} is defined in Equation (26), U_{MOVER} is defined in Equation (27), and l_i and u_i are defined in Equation (28).

Computational confidence interval

Theorem 1: Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ be a log-normal population with parameters μ_i and σ_i^2, where $i = 1, 2, \ldots, k$. For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$, let $X_{ij} = \log(Y_{ij})$ be the normal distribution with mean μ_i and variance σ_i^2. The maximum likelihood estimators of μ_i and θ given by Equation (3) under $\theta_1 = \theta_2 = \ldots = \theta_k = \theta$ are given by

$$\hat{\mu}_i = \bar{X}_i \qquad (30)$$

and

$$\sum_{i=1}^{k} \frac{n_i \hat{\sigma}_i^2}{\log(\theta^2 + 1)} - \sum_{i=1}^{k} n_i = 0. \qquad (31)$$

Proof: The log-likelihood function of normal distribution with parameters μ_i and θ is given by

$$\ln L = -\frac{1}{2}\sum_{i=1}^{k} n_i \log\left(2\pi \log\left(\theta^2+1\right)\right) - \sum_{i=1}^{k} n_i \log\left(Y_{ij}\right) - \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{\left(\log\left(Y_{ij}\right) - \mu_i\right)^2}{2\log\left(\theta^2+1\right)}.$$

Differentiating the $\ln L$ with respect to μ_i and θ, respectively, the maximum likelihood estimators of μ_i and θ are given by

$$\hat{\mu}_i = \bar{X}_i$$

and

$$\sum_{i=1}^{k} \frac{n_i \hat{\sigma}_i^2}{\log(\theta^2 + 1)} - \sum_{i=1}^{k} n_i = 0.$$

189 Hence, Theorem 1 is proved.

According to Pal et al. (2007), the computational approach uses the maximum likelihood estimates (MLEs). The common coefficient of variation based on maximum likelihood estimator is defined by

$$\hat{\theta}_{ML} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{\widehat{Var}(\hat{\theta}_i)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\widehat{Var}(\hat{\theta}_i)}, \tag{32}$$

190 where $\hat{\theta}_i = \sqrt{\exp\left(S_i^2\right) - 1}$ and $\widehat{Var}\left(\hat{\theta}_i\right)$ is defined in Equation (4) with σ_i replaced by s_i.

191 The computational approach is to obtain the restricted maximum likelihood estimates (RMLEs) of
192 parameters. The maximum likelihood estimators of μ_i and θ under $\theta_1 = \theta_2 = \ldots = \theta_k = \theta$ provide the
193 RMLEs of these parameters.

194 Then the RMLE of μ_i is defined by $\hat{\mu}_{i(RML)} = \bar{X}_i$. The RMLE of θ obtained iteratively from Equa-
195 tion (31) by using bisection method. The θ converge to the RMLE denoted as $\hat{\theta}_{RML}$.

196 Data replication from $f(x|\hat{\mu}_{i(RML)}, \hat{\theta}_{RML})$ is used to construct the confidence interval based on com-
197 putational approach. Let artificial sample $X_{i(RML)} = \left(X_{i1(RML)}, X_{i2(RML)}, \ldots, X_{in_i(RML)}\right)$ be the normal
198 distribution with mean $\hat{\mu}_{i(RML)}$ and variance $\hat{\sigma}^2_{i(RML)}$. Let $\bar{X}_{i(RML)}$ and $S^2_{i(RML)}$ be the mean and variance of
199 the log-transformed sample from a log-normal distribution for the i-th artificial sample and let $\bar{x}_{i(RML)}$ and
200 $s^2_{i(RML)}$ be observed sample mean and observed sample variance, respectively.

Therefore, the common coefficient of variation based on k individual samples is defined by

$$\hat{\theta}_{RML} = \sum_{i=1}^{k} \frac{\hat{\theta}_{i(RML)}}{\widehat{Var}(\hat{\theta}_{i(RML)})} \Bigg/ \sum_{i=1}^{k} \frac{1}{\widehat{Var}(\hat{\theta}_{i(RML)})}, \tag{33}$$

201 where $\hat{\theta}_{i(RML)} = \sqrt{\exp\left(S^2_{i(RML)}\right) - 1}$.

Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common coefficient of variation θ based on computational approach is

$$CI_{CA} = [L_{CA}, U_{CA}] = [\hat{\theta}_{RML}(\alpha/2), \hat{\theta}_{RML}(1-\alpha/2)], \tag{34}$$

202 where $\hat{\theta}_{RML}(\alpha/2)$ and $\hat{\theta}_{RML}(1-\alpha/2)$ denote the $(\alpha/2)$-th and $(1-\alpha/2)$-th percentiles of $\hat{\theta}_{RML}$, respec-
203 tively.

204 The following algorithm is used to construct the computational confidence interval:

205 **Algorithm 2**
206 For a given $\bar{x}_i$, s_i^2, and θ, where $i = 1, 2, \ldots, k$
207 Compute $\hat{\mu}_{i(RML)}$ and $\hat{\theta}_{RML}$ from Equation (30)-Equation (31)
208 For $g = 1$ to m
209 Generate $x_{ij(RML)}$ from $N\left(\hat{\mu}_{i(RML)}, \sqrt{\log\left(\hat{\theta}^2_{RML} + 1\right)}\right)$
210 Compute $\bar{x}_{i(RML)}$ and $s^2_{i(RML)}$
211 Compute $\hat{\theta}_{RML}$ from Equation (33)
212 End g loop
213 Compute $\hat{\theta}_{RML}(\alpha/2)$ and $\hat{\theta}_{RML}(1-\alpha/2)$ from Equation (34)
214

215 **Bayesian confidence interval**

The FGCI approach, MOVER approach, and computational approach are the classical approach. The classical approach and the Bayesian approach are fundamentally different. In the classical approach, the parameter of interest θ is unknown, but it is fixed. In the Bayesian approach, the parameter is considered to be a quantity. The variation of the quantity is described by the prior distribution. Bayes (1763) introduced that Bayesian approach uses Bayes' theorem to update probabilities. Bayes' theorem describes the conditional probability of an event based on data. The data is prior information or beliefs about the event. The posterior distribution is combination of the likelihood function and the prior distribution. The Bayesian confidence interval is constructed based on the posterior distribution. The posterior distribution

is a conditional distribution which is based on the observed values of the sample. The posterior distribution is used to make statements about the parameter. The parameter is considered a random quantity. The conditional posterior distribution for μ_i given σ_i^2 and x_i is the normal distribution with mean $\hat{\mu}_i$ and variance σ_i^2/n_i. The distribution is defined by

$$\mu_i|\sigma_i^2, x_i \sim N(\hat{\mu}_i, \sigma_i^2/n_i). \tag{35}$$

The posterior distribution for σ_i^2 is inverse gamma distribution. It is defined by

$$\sigma_i^2|x_i \sim IG((n_i-1)/2, (n_i-1)s_i^2/2). \tag{36}$$

The posterior distribution of coefficient of variation of log-normal distribution is

$$\hat{\theta}_i = \sqrt{\exp\left(\sigma_i^2\right)-1}, \tag{37}$$

where σ_i^2 is defined in Equation (36).

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

$$Var(\hat{\theta}_i) \approx \frac{\sigma_i^4(\exp(2\sigma_i^2))}{2(n_i-1)(\exp(\sigma_i^2)-1)}, \tag{38}$$

where σ_i^2 is defined in Equation (36).

The common coefficient of variation of log-normal distribution based on k individual samples which the parameter of interest defined by

$$\hat{\theta}_{BS} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{Var(\hat{\theta}_i)} \bigg/ \sum_{i=1}^{k} \frac{1}{Var(\hat{\theta}_i)}, \tag{39}$$

where $\hat{\theta}_i$ is defined in Equation (37) and $Var(\hat{\theta}_i)$ is defined in Equation (38).

Gelman et al. (2013) introduced the highest posterior density interval to construct the Bayesian confidence interval. Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common coefficient of variation θ based on Bayesian approach is

$$CI_{BS} = [L_{BS}, U_{BS}], \tag{40}$$

where L_{BS} and U_{BS} are the lower limit and the upper limit of the shortest $100(1-\alpha)\%$ highest posterior density interval of $\hat{\theta}_{BS}$, respectively.

The following algorithm is used to construct the Bayesian confidence interval:

Algorithm 3

For a given $\bar{x}_i$ and s_i^2, where $i = 1, 2, \ldots, k$

For $g = 1$ to m

Generate $\mu_i|\sigma_i^2, x_i \sim N(\hat{\mu}_i, \sigma_i^2/n_i)$

Generate $\sigma_i^2|x_i \sim IG((n_i-1)/2, (n_i-1)s_i^2/2)$

Compute θ_i from Equation (37)

Compute $Var(\hat{\theta}_i)$ from Equation (38)

Compute θ_{BS} from Equation (39)

End g loop

Compute L_{BS} and U_{BS}

RESULTS

A simulation study was performed to evaluate the coverage probabilities and average lengths of the FGCI (CI_{FGCI}), MOVER (CI_{MOVER}), computational (CI_{CA}), and Bayesian confidence intervals (CI_{BS}). The confidence intervals were compared by measuring their coverage probabilities and average lengths and, in each case, the one with a coverage probability closest to the nominal confidence level $(1-\alpha)$ and with the shortest average length was chosen as the most appropriate.

In this simulation study, the nominal confidence level was chosen as 0.95. The sample cases used were $k = 3$ and $k = 6$ with sample sizes $n_1, n_2, \ldots, n_k$, as in Tables 1 and 2. Following Tian and Wu (2007) and Ng (2014), the coefficient of variation of log-normal distribution, which equals $\sqrt{\exp(\sigma^2) - 1}$, is chosen in the range from 0.05 - 2.00. Since the coefficient of variation of a log-normal distribution is independent of μ, the population means of the normal data within each sample were given the same value $\mu_1 = \mu_2 = \ldots = \mu_k = \mu = 1$ to simplify matters, and the population standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_k$ are as in Tables 1 and 2. For each parameter setting, 5000 random samples were generated by applying Algorithm 4, and thus $1000R_\theta$, $1000\hat{\theta}_{RML}$, and $1000\theta_{BS}$ were simulated by applying Algorithms 1, 2, and 3, respectively, for each of the random samples.

The following algorithm is used to estimate the coverage probability and average length:

Algorithm 4

For a given $(n_1, n_2, \ldots, n_k)$, $(\mu_1, \mu_2, \ldots, \mu_k)$, $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ and θ

For $h = 1$ to M

Generate x_{ij} from $N(\mu_i, \sigma_i^2)$, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$

Calculate $\bar{x}_i$ and s_i^2

Construct $[L_{FGCI(h)}, U_{FGCI(h)}]$

Construct $[L_{MOVER(h)}, U_{MOVER(h)}]$

Construct $[L_{CA(h)}, U_{CA(h)}]$

Construct $[L_{BS(h)}, U_{BS(h)}]$

Record whether or not all the values of θ fall in their corresponding confidence intervals

Compute $U_{(h)} - L_{(h)}$

End h loop

Compute the coverage probability and the average length for each confidence interval

Tables 1 and 2 report the coverage probabilities and average lengths for $k = 3$ and $k = 6$, respectively. From Table 1, the simulation results indicate that for all sample sizes, the FGCI approach provided the best coverage probabilities whereas the MOVER confidence interval attained coverage probabilities under the nominal confidence level of 0.95. Furthermore, the coverage probabilities of the MOVER confidence interval decreased when the sample sizes increased. The computational confidence interval achieved coverage probabilities under the nominal confidence level of 0.95, which became closer to it as the sample sizes increased. The coverage probabilities of the Bayesian confidence interval performed well when the population standard deviations were (0.05,0.10,0.15) whereas they were less than the nominal confidence level of 0.95 when the population standard deviations were (0.50,1.00,1.00).

From Table 2, it can be seen that the coverage probabilities of the FGCI and the computational confidence interval were less than the nominal confidence level of 0.95 when the sample sizes were small. For large sample sizes, the coverage probabilities of the FGCI and the computational confidence interval were close to the nominal confidence level of 0.95, with those of the FGCIs being closer. The coverage probabilities of the MOVER confidence interval were greater than the nominal confidence level of 0.95 for small sample sizes and became close to 1.00 when the sample sizes increased, thereby showing conservative behavior when the sample case (k) was large and the sample sizes were large. Therefore, the MOVER approach can be considered as an alternative to estimate the confidence interval for the common coefficient of variation of log-normal distributions when the sample case ($k = 6$) is large and the sample sizes are small. The coverage probabilities of the Bayesian confidence interval were less than the nominal confidence level of 0.95 when the sample sizes were small and were close to the nominal confidence level of 0.95 when the sample sizes were large. The average lengths of the Bayesian confidence interval were shorter than those of the others.

As the sample case (k) increased, the coverage probabilities of the FGCI and the computational confidence interval tended to decrease because the common coefficient of variation θ is based on the variance of the coefficient of variation $\hat{\theta}_i$ for the k individual samples. Herein, we present only the results of $\mu = 1$ because they tended toward the same direction regardless of the value of μ. In all cases, the coverage probabilities were affected by large σ values because the coefficient of variation of log-normal distributions $\theta = \sqrt{\exp(\sigma^2) - 1}$ depends on parameter σ only.

Table 1. The coverage probabilities and average lengths of 95% two-sided confidence intervals for the common coefficient of variation of several log-normal populations: 3 sample cases.

(n_1, n_2, n_3)	$(\sigma_1, \sigma_2, \sigma_3)$	Coverage probability (Average length)			
		CI_{FGCI}	CI_{MOVER}	CI_{CA}	CI_{BS}
(30,30,30)	(0.05,0.10,0.15)	0.9514	0.9238	0.9188	0.9400
		(0.0317)	(0.0306)	(0.0305)	(0.0309)
	(0.50,1.00,1.00)	0.9496	0.9150	0.9246)	0.9394
		(0.3922)	(0.3672)	(0.3628	(0.3781)
(50,50,50)	(0.05,0.10,0.15)	0.9492	0.9234	0.9244	0.9408
		(0.0246)	(0.0230)	(0.0240)	(0.0241)
	(0.50,1.00,1.00)	0.9504	0.9054	0.9372	0.9442
		(0.3060)	(0.2707)	(0.2909)	(0.2976)
(30,50,100)	(0.05,0.10,0.15)	0.9530	0.9160	0.9200	0.9426
		(0.0331)	(0.0312)	(0.0329)	(0.0325)
	(0.50,1.00,1.00)	0.9498	0.8808	0.9186	0.9360
		(0.4259)	(0.3831)	(0.4055)	(0.4147)
(50,100,200)	(0.05,0.10,0.15)	0.9528	0.9222	0.9358	0.9430
		(0.0251)	(0.0234)	(0.0252)	(0.0247)
	(0.50,1.00,1.00)	0.9492	0.8722	0.9314	0.9410
		(0.3357)	(0.2875)	(0.3272)	(0.3291)
(100,100,100)	(0.05,0.10,0.15)	0.9502	0.9214	0.9388	0.9442
		(0.0174)	(0.0159)	(0.0172)	(0.0171)
	(0.50,1.00,1.00)	0.9478	0.8952	0.9398	0.9394
		(0.2176)	(0.1852)	(0.2117)	(0.2129)
(200,200,200)	(0.05,0.10,0.15)	0.9446	0.9178	0.9396	0.9404
		(0.0123)	(0.0111)	(0.0122)	(0.0121)
	(0.50,1.00,1.00)	0.9488	0.8944	0.9456	0.9434
		(0.1542)	(0.1288)	(0.1520)	(0.1513)
(500,500,500)	(0.05,0.10,0.15)	0.9482	0.9200	0.9444	0.9420
		(0.0078)	(0.0070)	(0.0078)	(0.0077)
	(0.50,1.00,1.00)	0.9480	0.8940	0.9470	0.9430
		(0.0976)	(0.0807)	(0.0971)	(0.0961)
(1000,1000,1000)	(0.05,0.10,0.15)	0.9428	0.9162	0.9446	0.9390
		(0.0055)	(0.0049)	(0.0055)	(0.0054)
	(0.50,1.00,1.00)	0.9574	0.9020	0.9552	0.9524
		(0.0691)	(0.0569)	(0.0689)	(0.0681)

An empirical application

The rainfall has been the seasonal problem in Thailand. Rainy season in Thailand is between May and October. The daily rainfall appears on 17 June 2020. Real data example of rainfall data is used to illustrate the FGCI, MOVER, computational, and Bayesian approaches. All data were reported by the Thai Meteorological Department (https://www.tmd.go.th/climate/climate.php).

The rainfall data on 17 June 2020 in Northern, Northeastern, Central, Eastern, and Southern regions are reported in Dataset 1, and histogram and normal QQ-plots are presented in Figures 1 and 2, respectively. The data sets consist of 30 measurements in Northern region, 31 measurements in Northeastern region, 21 measurements in Central region, 16 measurements in Eastern region, and 27 measurements in Southern region. The statistics is summarized in Table 3. The Shapiro-Wilk normality test is used to check the assumption that the log-data is normal distribution. The Shapiro-Wilk normality test with p-values 0.2176, 0.4981, 0.0009, 0.0128, and 0.2127 for Northern, Northeastern, Central, Eastern, and Southern regions, respectively. From p-values, the results show that the rainfall of the three regions follow log-normal distributions such as Northern, Northeastern, and Southern regions. The data of Northern, Northeastern, and Southern regions were used to construct the confidence interval for the common coefficient of variation based on the four approaches. The true common coefficient of variation was 1.2084. The point estimators of the common coefficient of variation based on the FGCI, MOVER, CA, and Bayesian approaches

308   were 1.2776,1.1486, 1.2273, and 1.2669. The FGCI, MOVER, computational, and Bayesian confidence
309   intervals were [0.8380,2.0301] with an interval length of 1.1921, [0.8460,1.8481] with an interval length
310   of 1.0021, [0.7745,1.8663] with an interval length of 1.0918, and [0.7991,1.7704] with an interval length
311   of 0.9713, respectively. Hence, the length of the Bayesian confidence interval was shorter than those of
312   the others, and so was more accurate.
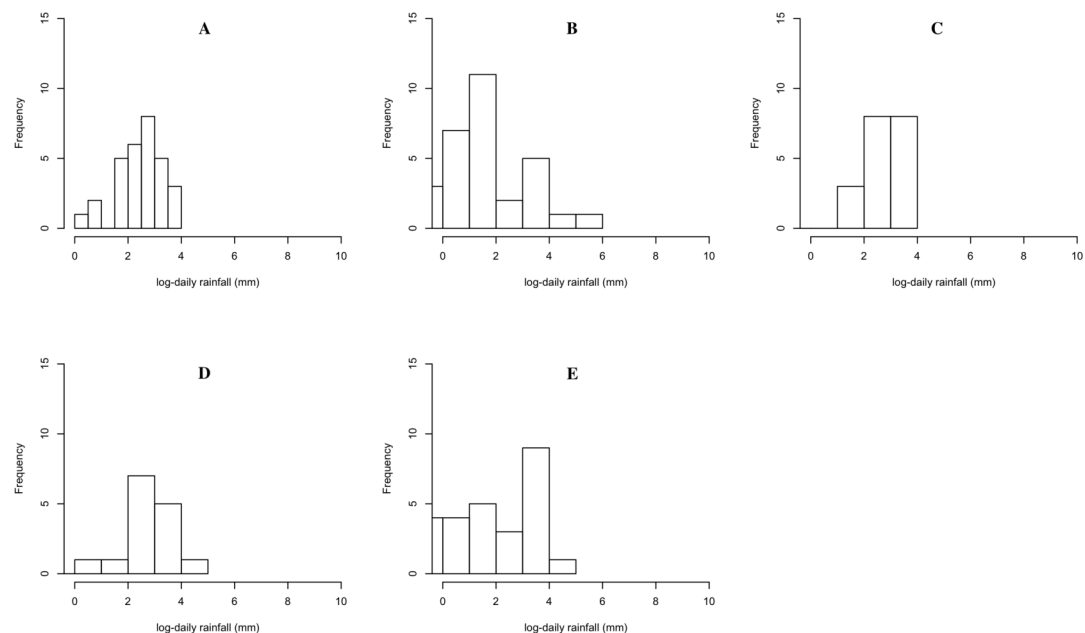
Figure 1. Histogram of rainfall data in (A) Northern (B) Northeastern (C) Central (D) Eastern (E) Southern regions.

DISCUSSION

313

314   As a limitation of this study, the coefficient of variation of log-normal distribution can be computed directly
315   from log-normal data and the data transformed using the log function. The coefficients of variation based
316   directly on log-normal data make it easy to construct the confidence interval for the common coefficient
317   of variation of log-normal distributions because the coefficient of variation of a log-normal distribution is
318   based on σ^2 only. However, for a greater number of samples, i.e. $n = 500$ and a large σ^2, i.e. $\sigma^2 = 10$,
319   the coverage probabilities of all of the proposed confidence intervals cannot be computed (the simulation
320   results are not reported here). The coefficients of variation based on transformed data using the log
321   function make it more difficult to estimate the confidence interval for the common coefficient of variation
322   of log-normal distributions because the coefficient of variation of a normal distribution is based on both μ
323   and σ^2.
324   The MOVER approach makes it possible to compute the confidence interval online because this
325   approach requires a simple formula to construct the confidence interval. Since the FGCI, computational,
326   and Bayesian approaches are based on simulation techniques, it is not possible to compute the confidence
327   intervals for them online.
328   As a final note, the results of the computational approach did not perform well for the confidence
329   interval estimation for the common coefficient of variation of log-normal distributions. However, Thangjai
330   et al. (2020b) reported that it performs well for constructing the confidence interval for the common
331   coefficient of variation of normal distributions when the sample case is large.
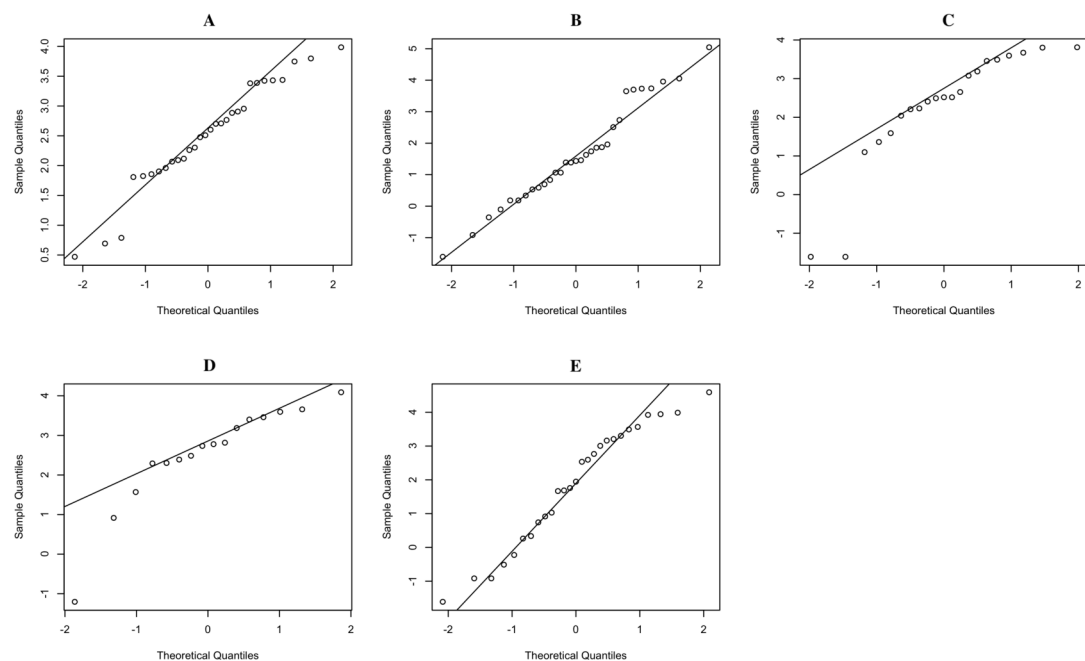
Figure 2. QQ-plot of log- rainfall data in (A) Northern (B) Northeastern (C) Central (D) Eastern (E) Southern regions.

CONCLUSIONS

The results in Tables 1 and 2 indicate that the FGCI approach provided much better confidence interval estimates than the other approaches in terms of coverage probability for almost all sample cases (k) and sample sizes (n), except that the MOVER approach provided the best confidence interval estimates when the sample case was equal to six populations ($k = 6$) and the sample sizes were small ($n_i < 50$). Moreover, the FGCI approach was the best for constructing the confidence interval for all sample sizes (n) when the sample case is small ($k = 3$), for which the coverage probability of the FGCI approach was stable around the nominal confidence level of 0.95. For large sample cases ($k = 6$), the FGCI approach performed well for confidence interval estimation when the sample sizes were large. Furthermore, the FGPQ for the FGCI approach is not dependent on the population variances (σ_i). The results are similar to those of Hannig et al. (2006b), Chang and Huang (2009), Kharrati-Kopaei et al. (2013), and Thangjai et al. (2016). Note that the GCIs based on Ng (2014) had coverage probabilities close to 1.00 when the sample size was small and close to the nominal confidence level of 0.95 when the sample size was large. Therefore, the GCIs based on Ng (2014) are rather conservative for small sample sizes.

REFERENCES

Bayes T. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**:370–418.

Casella G, Berger RL. 2002. *Statistical Inference*. California: Duxbury Press.

Chang YP, Huang WT. 2009. Simultaneous fiducial generalized confidence intervals for all pairwise comparisons of means. *International Journal of Information and Management Sciences* **20**:459–467.

Donner A, Zou GY. 2012. Closed-form confidence intervals for function of the normal standard deviation. *Statistical Methods in Medical Research* **21**:347–359.

Fung WK, Tsang TS. 1998. A simulation study comparing tests for the equality of coefficients of variations. *Statistics in Medicine* **17**:2003–2014.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013. *Bayesian data analysis*.

357 Florida: Chapman and Hall/CRC.
358 **Gokpinar F, Gokpinar E. 2015.** A computational approach for testing equality of coefficients of varia-
359 tion in k normal populations. *Hacettepe Journal of Mathematics and Statistics* **44**:1197–1213.
360 **Gokpinar F, Gokpinar E. 2017.** Testing the equality of several lognormal means based on a computa-
361 tional approach. *Communications in Statistics - Simulation and Computation* **46**:1998–2010.
362 **Graybill FA, Deal RB. 1959.** Combining unbiased estimators. *Biometrics* **15**:543–550.
363 **Gupta RC, Li X. 2005.** Statistical inference for the common mean of two log-normal distributions and
364 some applications in reliability. *Computational Statistics and Data Analysis* **50**:3141–3164.
365 **Gupta RC, Ma S. 1996.** Testing the equality of the coefficients of variation in *k* normal populations.
366 *Communications in Statistics - Theory and Methods* **25**:115–132.
367 **Gutiérrez R, Rico N, Román P, Torres F. 2007.** Approximate and generalized confidence bands for the
368 mean and mode functions of the lognormal diffusion process. *Computational Statistics and Data Analysis*
369 **51**:4038–4053.
370 **Gutiérrez R, Román P, Romero D, Torres F. 2003.** Forecasting for the univariate lognormal diffusion
371 process with exogenous factors. *Cybernetics and Systems* **34**:709–724.
372 **Hannig J, Iyer H, Patterson P. 2006a.** Fiducial generalized confidence intervals. *Journal of the Ameri-
373 can Statistical Association* **101**:254–269.
374 **Hannig J, Lidong E, Abdel-Karim A, Iyer H. 2006b.** Simultaneous fiducial generalized confidence
375 intervals for ratios of means of lognormal distributions. *Austrian Journal of Statistics* **35**:261–269.
376 **Hasan MS, Krishnamoorthy K. 2017.** Improved confidence intervals for the ratio of coefficients of
377 variation of two lognormal distributions. *Journal of Statistical Theory and Applications* **16**:345–353.
378 **Jafari AA, Abdollahnezhad K. 2015.** Inferences on the means of two log-Normal distributions: A
379 computational approach test. *Communications in Statistics - Simulation and Computation* **44**:1659–1672.
380 **Jafari AA, Kazemi MR. 2017.** Computational approach test for inference about several correlation
381 coefficients: Equality and common. *Communications in Statistics - Simulation and Computation* **46**:2043–
382 2056.
383 **Joulious SA, Debarnot CAM. 2000.** Why are pharmacokinetics data summarize by arithmetic means?.
384 *Journal of Biopharmaceutical Statistics* **10**:55–71.
385 **Kharrati-Kopaei M, Malekzadeh A, Sadooghi-Alvandi M. 2013.** Simultaneous fiducial generalized
386 confidence intervals for the successive differences of exponential location parameters under heteroscedas-
387 ticity. *Statistics and Probability Letters* **83**:1547–1552.
388 **Kelley K. 2007.** Sample size planning for the coefficient of variation from the accuracy in parameter
389 estimation approach. *Behavior Research Methods* **39**:755–766.
390 **Krishnamoorthy K, Lu Y. 2003.** Inference on the common means of several normal populations based
391 on the generalized variable method. *Biometrics* **59**:237–247.
392 **Krishnamoorthy K, Mathew T. 2003.** Inferences on the means of lognormal distributions using gen-
393 eralized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*
394 **115**:103–121.
395 **Krishnamoorthy K, Oral E. 2017.** Standardized likelihood ratio test for comparing several log-normal
396 means and confidence interval for the common mean. *Statistical Methods in Medical Research* **26**:2919–
397 2937.
398 **Land CE. 1988.** Hypothesis tests and interval estimates. In Crow, E.L. and Shimizu, K. (Eds.). *Log-
399 normal distributions: Theory and applications*. New York: Marcel Dekker.
400 **Miller EC, Karson MJ. 1977.** Testing equality of two coefficients of variation. *Proceedings of the
401 Business and Economics* **Section Part I**:278–283.
402 **Nam JM, Kwon D. 2017.** Inference on the ratio of two coefficients of variation of two lognormal
403 distributions. *Communications in Statistics-Theory and Methods* **46**: 8575–8587.
404 **Niwitpong S. 2013.** Confidence intervals for coefficient of variation of lognormal distribution with
405 restricted parameter space. *Applied Mathematical Sciences* **7**:3805–3810.
406 **Niwitpong, S. and Wongkhao, A. 2016.** Confidence intervals for the difference between inverse of
407 normal means. *Advances and Applications in Statistics* **48**:337–347.
408 **Ng CK. 2014.** Inference on the common coefficient of variation when populations are lognormal: A
409 simulation-based approach. *Journal of Statistics: Advances in Theory and Applications* **11**:117–134.
410 **Pal N, Lim WK, and Ling CH. 2007.** A computational approach to statistical inferences. *Journal of
411 Applied Probability and Statistics* **2**:13–35.

412 **Rumburg B, Alldredge R, Claiborn C. 2001.** Statistical distributions of particulate matter and the error
413 associated with sampling frequency. *Atmospheric Environment* **35**:2907–2920.
414 **Sangnawakij P, Niwitpong S. 2016.** Confidence intervals for coefficients of variation in two-parameter
415 exponential distributions. *Communications in Statistics - Simulation and Computation* **46**:6618–6630.
416 **Sangnawakij P, Niwitpong S, Niwitpong S. 2016.** Confidence intervals for the ratio of coefficients of
417 variation in the two-parameter exponential distributions. *Lecture Notes in Artificial Intelligence* **9978**:542–
418 551.
419 **Shen H, Brown L, Hui Z. 2006.** Efficient estimation of log-normal means with application to pharma-
420 cokinetics data. *Statistics in Medicine* **25**:3023–3038.
421 **Suwan S, Niwitpong S. 2013.** Estimated variance ratio confidence interval of nonnormal distributions.
422 *Far East Journal of Mathematical Sciences* **4**:339–350.
423 **Taylor DJ, Kupper LL, Muller KE. 2002.** Improved approximate confidence intervals for the mean of
424 a log-normal variable. *Statistics in Medicine* **21**:1443–1459.
425 **Tian L. 2005.** Inferences on the common coefficient of variation. *Statistics in Medicine* **24**:2213–2220.
426 **Tian L, Wu J. 2007.** Inferences on the common mean of several log-normal populations: The generalized
427 variable approach. *Biometrical Journal* **49**:944–951.
428 **Thangjai W, Niwitpong S, Niwitpong S. 2016.** Simultaneous fiducial generalized confidence intervals
429 for all differences of coefficients of variation of log-normal distributions. *Lecture Notes in Artificial*
430 *Intelligence* **9978**:552–561.
431 **Thangjai W, Niwitpong S, Niwitpong S. 2020a.** Simultaneous confidence intervals for all differences
432 of coefficients of variation of log-normal distributions. *Hacettepe Journal of Mathematics and Statistics*
433 **48**:1505–1521.
434 **Thangjai W, Niwitpong S, Niwitpong S. 2020b.** Adjusted generalized confidence intervals for the
435 common coefficient of variation of several normal populations. *Communications in Statistics-Simulation*
436 *and Computation* **49**:194–206.
437 **Tsim YL, Yip SP, Tsang, KS, Li KF, Wong HF. 1991.** Haematogoy and Serology. In *Annual Report,*
438 *Honk Kong Medical Technology Association Quality Assurance Programme*: 25–40.
439 **Weerahandi S. 1993.** Generalized confidence intervals. *Journal of American Statistical Association*
440 **88**:899–905.
441 **Ye RD, Ma TF, Wang SG. 2010.** Inferences on the common mean of several inverse Gaussian popula-
442 tions. *Computational Statistics and Data Analysis* **54**:906–915.
443 **Zou GY, Donner A. 2008.** Construction of confidence limits about effect measures: A general approach.
444 *Statistics in Medicine* **27**:1693–1702.
445 **Zou GY, Taleban J, Hao CY. 2009.** Confidence interval estimation for lognormal data with application
446 to health economics. *Computation Statistics and Data Analysis* **53**: 3755–3764.
447 **Zhou XH. 1998.** Estimation of the log-normal mean. *Statistics in Medicine* **17**:2251–2264.
448 **Zhou XH, Gao SJ. 1997.** Confidence intervals for the log-normal mean. *Statistics in Medicine* **16**:783–
449 790.
450

Table 2. The coverage probabilities and average lengths of 95% two-sided confidence intervals for the common coefficient of variation of several log-normal populations: 6 sample cases.

| $(n_1,n_2,n_3,n_4,n_5,n_6)$ | $(\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5,\sigma_6)$ | Coverage probability (Average length) | | | |
		CI_{FGCI}	CI_{MOVER}	CI_{CA}	CI_{BS}
(30,30,30,30,30,30)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9432	0.9818	0.8812	0.9328
		(0.0226)	(0.0298)	(0.0227)	(0.0222)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9218	0.9838	0.8418	0.8982
		(0.2099)	(0.3278)	(0.2108)	(0.2054)
(30,50,100,30,50,100)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9400	0.9800	0.8928	0.9314
		(0.0199)	(0.0247)	(0.0202)	(0.0196)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9350	0.9830	0.8992	0.9224
		(0.1490)	(0.2033)	(0.1505)	(0.1461)
(50,50,50,50,50,50)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9392	0.9834	0.9060	0.9342
		(0.0175)	(0.0227)	(0.0175)	(0.0172)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9290	0.9890	0.8812	0.9182
		(0.1628)	(0.2503)	(0.1638)	(0.1598)
(30,30,50,50,100,100)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9354	0.9808	0.8636	0.9218
		(0.0239)	(0.0305)	(0.0245)	(0.0235)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9218	0.9800	0.8596	0.9048
		(0.2128)	(0.2931)	(0.2148)	(0.2086)
(50,50,100,100,200,200)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9432	0.9860	0.9028	0.9360
		(0.0180)	(0.0231)	(0.0184)	(0.0178)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9354	0.9822	0.9008	0.9218
		(0.1598)	(0.2103)	(0.1611)	(0.1571)
(30,30,50,100,100,200)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9398	0.9800	0.8678	0.9312
		(0.0226)	(0.0283)	(0.0236)	(0.0223)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9284	0.9788	0.8618	0.9100
		(0.2257)	(0.2980)	(0.2291)	(0.2216)
(100,100,100,100,100,100)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9480	0.9830	0.9282	0.9406
		(0.0123)	(0.0158)	(0.0124)	(0.0122)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9362	0.9952	0.9166	0.9256
		(0.1149)	(0.1747)	(0.1154)	(0.1131)
(200,200,200,200,200,200)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9512	0.9872	0.9410	0.9428
		(0.0087)	(0.0111)	(0.0087)	(0.0086)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9418	0.9968	0.9322	0.9364
		(0.0813)	(0.1231)	(0.0816)	(0.0801)
(500,500,500,500,500,500)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9510	0.9876	0.9440	0.9464
		(0.0055)	(0.0070)	(0.0055)	(0.0054)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9520	0.9960	0.9494	0.9494
		(0.0514)	(0.0776)	(0.0514)	(0.0506)
(1000,1000,1000,1000,1000,1000)	(0.05,0.05,0.10,0.10,0.15,0.15)	0.9520	0.9890	0.9496	0.9490
		(0.0039)	(0.0049)	(0.0039)	(0.0038)
	(0.50,0.50,0.50,1.00,1.00,1.00)	0.9520	0.9974	0.9488	0.9470
		(0.0363)	(0.0548)	(0.0363)	(0.0358)

Table 3. Sample statistics of Northern, Northeastern, Central, Eastern, and Southern regions (mm)

Statistics	Northern region	Northeastern region	Central region	Eastern region	Southern region
n	30	31	21	16	27
$\bar{y}$	17.1200	16.8677	17.9333	19.9688	18.3852
s_Y	13.5250	30.9424	14.7627	15.8678	23.2974
$\bar{x}$	2.5082	1.6238	2.2851	2.5292	1.8608
s_X	0.8945	1.6007	1.5163	1.2827	1.7549
$\hat{\theta}$	1.1072	3.4592	2.9942	2.0450	4.5550