Quantifying uncertainty of emission estimates in National Greenhouse Gas Inventories using bootstrap confidence intervals

Lee-Ing Tong*, Chih-Wei Chang, Shin-En Jin, R. Saminathan

Dept. of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan

A B S T R A C T

Greenhouse gas (GHG) emissions have exacerbated global warming, and consequently are the focus of worldwide reduction efforts. Reducing emissions involves accurately estimating GHG emissions and the uncertainty associated with such estimates. The uncertainty of GHG emission estimates is often assessed using the 95% confidence interval. Given a small sample size and non-normal distribution of the underlying population, the uncertainty estimate obtained using the 95% confidence interval may lead to significant bias. Bootstrap confidence interval is an effective means of reducing bias. This work presents a procedure for estimating the uncertainty of GHG emission estimation using bootstrap confidence intervals. Numerical simulation is performed for GHG emission estimates under three distributions (namely normal, log-normal and uniform) to find the 95% confidence intervals and bootstrap confidence intervals. Finally, the accuracy and sensitivity of the uncertainty of various interval estimations are examined by comparing the coverage performance, interval mean and interval standard deviation. Simulation results indicate that the bootstrap intervals are more applicable than the 95% confidence interval given non-normal dataset and small sample size. Moreover, when sample size *n* is less than 30, the bootstrap confidence interval has a smaller interval length with a smaller deviation than that of the classical 95% confidence interval regardless of whether the data distribution is normal or non-normal. This study recommends a sample size greater than or equal to 9 for estimating the uncertainty of emission estimates. When the sample size *n* exceeds 30, either the normality-based 95% confidence interval or bootstrap confidence intervals may be used regardless of whether the data distribution is normal or non-normal. A case study of carbon stock from Taiwan demonstrates the feasibility of the proposed procedure.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Global warming caused by anthropogenic greenhouse gas (GHG) emissions has contributed significantly to global climate change. The Intergovernmental Panel on Climate Change (IPCC) predicts global temperatures will rise between 1.8 °C and 4.0 °C by 2100. The United Nations Framework Convention on Climate Change (UNFCCC) was held in 1992 to control GHG emissions and to reduce its influence on climate change. The Kyoto Protocol was developed under the UNFCCC to address global warming in 1997. The Kyoto Protocol mandates industrialized countries to reduce GHG emissions by 5.2% relative to 1990 emission levels by 2012.

Variability refers to the heterogeneity of values within a population and is an inherent property of either the system or of nature, and not of the analyst. However, uncertainty refers to lack of knowledge regarding the true value of a quantity. Uncertainty depends on the state of analyst knowledge, which in turn depends on the quality and quantity of applicable data, as well as knowledge of underlying processes and inference methods (IPCC, 2006; Zheng and Frey, 2004). Reducing emissions initially involves accurately estimating GHG emissions and associated uncertainties. Uncertainty of emission estimates is an essential component of a complete and transparent emission inventory (IPCC, 2000). The IPCC has developed and refined guidance on quantifying uncertainty in National GHG emission inventories, including Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories (IPCC, 2000), Good Practice Guidance for Land Use Change and Forestry (IPCC, 2003), and IPCC Guidelines for National Greenhouse Inventories (IPCC, 2006). These guidelines defined the uncertainty range to express uncertainties regarding the estimated GHG emissions. If the ranges of uncertainty are small, then analytical methods can achieve acceptable accuracy (IPCC,

2006). However, if the ranges of uncertainty are large, analytical methods can produce errors. The bootstrap simulation, an alternative to the conventional approach, was mentioned in Section 3.2.2.2 of IPCC (2006) when the relative range of uncertainty or the standard deviation of the mean is large. Uncertainties in inventories arise through at least three different processes (IPCC, 2000):

1. Uncertainties arising from definitions (e.g. incomplete meaning, unclear or incorrect definition of an emission or uptake);
2. Uncertainties arising from natural variability of the processes that produce an emission or uptake;
3. Uncertainties arising from the assessment methods including: (i) measurement uncertainties; (ii) sampling uncertainties; (iii) uncertainties from incompletely described reference data; and (iv) expert judgment uncertainties; (v) uncertainties from model or method description and equations.

Uncertainties regarding emission estimates in GHG owing to sampling error can be quantified in terms of their mean or other statistics by using the 95% confidence interval of the true emission (IPCC, 2006). The confidence interval is defined in terms of the 2.5th and 97.5th percentiles of the sampling distribution for uncertainty in a statistics (e.g. sample mean). The statistical literature outlines many probability density functions that frequently represent specific real world data. Some of these functions, such as normal, log-normal and uniform distributions, are considered for representing variability in activity data and emission factor data (IPCC, 2006).

Emission estimates are obtained by multiplying an appropriate emission factor by an activity data representing the extent of the emission generating activity. The approaches, such as classical statistical methods, expert judgment and the recommendations of IPCC, can be utilized to estimate uncertainty in the emission factor or activity data (IPCC, 2006). A sufficiently large dataset allows for direct use of classical statistical methods to estimate uncertainty. When empirical data are lacking or not considered fully representative for all causes of uncertainty, uncertainty must be estimated based on expert judgment or other methods recommended by IPCC. Uncertainty in a statistic (e.g. sample mean) attributable to random sampling error can be described using a sampling distribution. Sampling distributions are used to obtain confidence intervals for the parameters of a distribution. A confidence interval for a parameter is a measure of knowledge regarding the value of the parameter (Zheng and Frey, 2004). The 95% confidence interval for the true mean obtained using the classical statistical method relies on the central limit theory and large sample size. Therefore, when the number of emission data is relatively small and the emission data distribution is non-normal, the 95% confidence interval of the emission estimates obtained using the classical statistical method may result in significant bias.

An alternative method for quantifying uncertainty in the mean, or any statistic, owing to sampling error is the bootstrap simulation (Efron, 1979). Bootstrap simulation has been used to quantify uncertainty in emission factors and inventories for a variety of air pollutants, including GHGs, hazardous air pollutants (HAPs), and criteria pollutants. Examples of quantification of emissions factor uncertainty using bootstrap are Frey and Bammi (2002) and Frey and Li (2003). Examples of quantification of uncertainty in inventories based on bootstrap uncertainty estimates of emission factors are Frey and Zhao (2004) and Zhao and Frey (2004b). Furthermore, bootstrap simulation was used to develop an approach for quantifying both the uncertainty and variability of estimates of hazardous air pollutant (HAP) emissions from power plants (Frey and Rhodes, 1996). For instance, bootstrap simulation was applied to the cell

burner wall-fired dataset for estimating uncertainty in NO_2 emissions per Btu (Rhodes and Frey, 1997). Bootstrap simulation produces paired parameter estimates to represent regarding the distribution parameters. Similar work has been performed for highway vehicle emissions (Kini and Frey, 1997). A previous study demonstrated the feasibility of using bootstrap simulation to quantify uncertainty and variability for a selected example of NO_x emissions from coal-fired power plants (Frey and Zheng, 2002). Furthermore, the methods for bootstrap estimation of confidence intervals for emission factors with complete and censored data were applied to urban scale emission inventories (e.g., Frey and Zhao, 2004; Zhao and Frey, 2004b). Additionally, an unbiased method based on Maximum likelihood estimation (MLE) and bootstrap simulation can quantify the inter-unit variability and uncertainty in statistics such as mean for censored datasets of an air toxic emission factor (Zhao and Frey, 2004a). The effectiveness of this method was assessed by applying it to synthetic data with various degrees of censoring, sample sizes, coefficients of variation and detection limits, and by using various parametric distributions such as log-normal, gamma, and Weibull. According to those results, the unbiased method based on MLE and bootstrap simulation can robustly and reliably quantify the variability and uncertainty of censored datasets under variable conditions as outlined above.

Maximum likelihood estimation (MLE) and the bootstrap method are applied to extensive empirical emission factor data for combustion sources to quantify inter-unit variability and uncertainty in mean emissions for selected air toxics (Zhao and Frey, 2006). The largest range of uncertainty in the mean obtained for the external coal combustion benzene emission factor using 95% confidence interval was −93% to +411%. However, the detailed accuracy and sensitivity analyses for various bootstrap confidence intervals related to the uncertainty of estimates of GHG emission are rarely discussed. The accuracy and sensitivity analysis are fundamental and essential for numerical methods such as the bootstrap method and analysis ensures the estimation is reasonable.

This work aims to propose a methodology for quantifying the uncertainty of emission estimates using four bootstrap confidence intervals and to evaluate the performances of the proposed estimation methods by analyzing the accuracy and sensitivity of the confidence intervals of emission estimates. A case study of carbon stock from Taiwan is presented to illustrate the application of the proposed procedure. The robustness of bootstrap simulation can ensure reasonably accurate estimation of uncertainty.

2. Classical and bootstrap confidence intervals

To illustrate the utilization of the classical and bootstrap confidence intervals, the general definitions are presented in this section.

2.1. Classical confidence interval

A $(1 - \alpha)$ 100% confidence interval for an unknown parameter (e.g. population mean) is an interval calculated from the sample data, such that $(1 - \alpha)$ 100% of the intervals will enclose the true parameter value. For example, a 95% confidence interval is an interval with 0.95 probability to enclose the true parameter. Suppose that $\{x_1, x_2, \cdots, x_n\}$ is a random sample drawn from a normal population with unknown mean μ and unknown variance σ^2, then a $(1 - \alpha)$ 100% confidence interval for the true mean can be constructed as follows:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \tag{1}$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, n is the sample size, and $t_{\alpha/2}$ is the value corresponding to an upper-tail area of $\alpha/2$ in the t distribution with $n-1$ degrees of freedom. Relative to other methods of constructing a confidence interval (e.g. bootstrap simulation method), the 95% confidence interval for the mean obtained by Eq. (1) is referred as the "classical confidence interval" in this study.

When the sample size n is sufficiently large (i.e., $n \geq 30$), according to the Central Limit Theorem, sample mean ($\bar{x}$) is approximately normally distributed regardless of the distribution of the sampled population. For small samples, the sampling distribution of $\bar{x}$ depends on the particular form of the relative frequency distribution of the population being sampled, and the sample standard deviation s may not be a satisfactory approximation to the population standard deviation (Mendenhall and Sincich, 2007). As a result, utilizing classical method (i.e., Eq. (1)) to construct a 95% confidence interval for the population mean based on a small sample may lead to a bias. Meeden (1999) has verified that when sampling from a skewed population with small sample sizes, the usual confidence intervals for the mean have poor coverage properties.

It is notable that length of the confidence interval (i.e., the difference between the two endpoints of interval) indicates the precision of the interval estimates of the parameter from sample data. Moreover, in the definition of confidence interval, the confidence interval length is associated with confidence level. For fixed

sample size, higher confidence level generally produces a larger interval length, and shorter interval length usually has a lower confidence level. There is a trade-off between confidence level and precision of the estimate for the fixed sample size. On the other hand, for fixed confidence level, the interval length becomes shorter as the sample size increases. That is, larger sample sizes give shorter interval length or more precise estimates.

2.2. Bootstrap sampling and the bootstrap confidence interval

Efron (1979, 1982) introduced a non-parametric, but computationally intensive, estimation method called "bootstrap" for estimating the confidence interval of statistics. Bootstrap is a data-based simulation method for statistical inference. The main advantage of the non-parametric bootstrap method is that it does not rely on distributional assumptions. The bootstrap method thus can be used to estimate the sampling distribution of statistics (such as, sample mean) based on the assumption that the sample is representative of the population from which it is drawn, and that the observations are independently and identically distributed. Suppose $\{x_1, x_2, ..., x_n\}$ is a random sample of size n (parent sample) taken from a process. A bootstrap sample, denoted by $\{x_1^*, x_2^*, ..., x_n^*\}$, is then a sample of size n drawn (with replacement) from the original sample. Hence, there exist a total of n^n possible resamples. Bootstrap sampling is equivalent to sampling with replacement from the empirical probability distribution function. In practice,
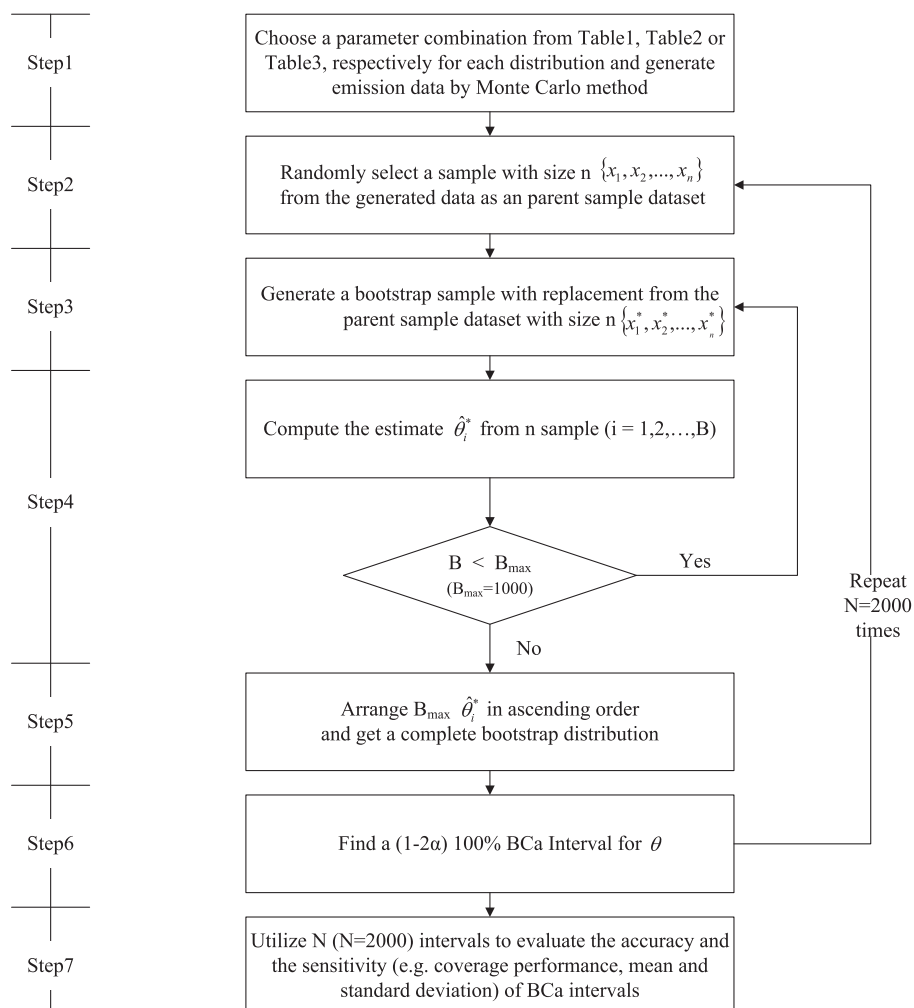

Fig. 1. Simulation procedure for BCa confidence interval of uncertainty of emission estimates using bootstrap method.

Table 1
Nine combinations of emission parameters with various sample sizes for a normal distribution.

Mean	Standard deviation	Sample size (n)
50, 60, 70	6, 9, 12	5, 6, 7,…, 30

usually only a random sample is drawn from the n^n possible resamples, the estimate is calculated for each of these, and the subsequent empirical distribution is referred to as the bootstrap distribution of the statistic. Efron and Tibshirani (1993) indicated that a minimum of approximately 1000 bootstrap resamples is sufficient to obtain accurate confidence interval estimates.

Suppose a random variable X is used to evaluate process performance. Although the distribution of X is unknown, the aim is to estimate some parameter θ that characterizes process performance, such as a mean emission factor. $\hat{\theta}$ can be estimated using the bootstrap sample. The estimate is represented by $\widehat{\theta}^*$ and is called the bootstrap estimate. The resampling procedure can be repeated numerous times, say, B times. Moreover, the B bootstrap estimates $\widehat{\theta}^*_1, \widehat{\theta}^*_2, …, \widehat{\theta}^*_B$ can be calculated from the resamples. Other studies on bootstrap methods include Efron and Gong (1983), Gunter (1991, 1992), Mooney and Duval (1993), and Young (1994). Efron and Tibshirani (1986) further developed four types of bootstrap confidence intervals; namely, the Standard Bootstrap (SB) confidence interval, Percentile Bootstrap (PB) confidence interval, Bias-Corrected Percentile Bootstrap (PCPB) confidence interval and Biased-Corrected Accelerated Percentile Bootstrap (BCa) confidence interval. The formulas used to calculate these intervals are detailed below:

1 Standard bootstrap (SB). From the B bootstrap estimates $\widehat{\theta}^*_1, \widehat{\theta}^*_2, …, \widehat{\theta}^*_B$ the sample mean $(\hat{\theta}^*)$ and standard deviation $(\hat{S}\hat{\theta}^*)$ of bootstrap estimates can be obtained as follows.

$$\hat{\theta}^* = \sum_{i=1}^{B} \hat{\theta}^*_i / B \tag{2}$$

$$\hat{S}\hat{\theta}^* = \sqrt{\left(\sum_{i=1}^{B} \left[\hat{\theta}^*_i - \hat{\theta}^* \right]^2 \right) \Big/ (B-1)} \tag{3}$$

where B denotes number of bootstrap resamples and $\hat{\theta}^*_i$ represents bootstrap estimates.

If the distribution of $\hat{\theta}^*$ is approximately normal, the $(1 - 2\alpha)$ 100% SB confidence interval for θ is $\hat{\theta}^* \pm z_\alpha \hat{S}\hat{\theta}^*$, where z_α is the 100 $(1 - 2\alpha)$th percentage point of the standard normal distribution.

2 Percentile bootstrap (PB). From the ordered collection of $\widehat{\theta}^*_1, \widehat{\theta}^*_2, …, \widehat{\theta}^*_B$, the $(1 - 2\alpha)$ 100% PB confidence interval for θ can be obtained as follows:

$$\left[\theta^*(\alpha B), \theta^*((1-\alpha)B) \right] \tag{4}$$

where $\hat{\theta}^*(i)$ is the ith value of ordered $\hat{\theta}^*_i$, $i = 1, 2, …, B$.

Table 2
Three combinations of emission parameters with various sample sizes for a uniform distribution.

(Lower Bound, Upper Bound)	Sample Size (n)
(45,75), (50,70), (55,65)	5, 6, 7,…, 30

Table 3
Nine combinations of emission parameters with various sample sizes for a log-normal distribution.

Mean	Standard deviation	Sample size (n)
50, 60, 70	15, 20, 25	5, 6, 7,…, 30

3 Biased-Corrected Percentile Bootstrap (BCPB). The bootstrap distribution may be biased. Consequently, the third approach is designed to correct this potential bias of the bootstrap distribution (see Efron, 1982 for a complete justification of the method). First the distribution of $\hat{\theta}^*(i)$ is used to calculate the probability,

$$p_0 = \Pr(\hat{\theta}^*(i) \le \hat{\theta}), \quad (i = 1, 2, …, B) \tag{5}$$

where $\hat{\theta}^*$ is the value of θ estimated from a random sample $\{x_1, x_2, …x_n\}$.

Second, the following quantities are calculated:

$$z_0 = \Phi^{-1}(p_0) \tag{6}$$

$$P_L = \Phi(2z_0 - z_\alpha) \tag{7}$$

$$P_U = \Phi(2z_0 + z_\alpha) \tag{8}$$

where $\Phi(.)$ denotes the cumulative standard normal distribution function.

Then the 100 $(1 - 2\alpha)$% BCPB confidence interval for θ is obtained as $(\hat{\theta}^*(P_L \times B), \hat{\theta}^*(P_U \times B))$.

4. Bias-Corrected and Accelerated (BCa) bootstrap. The P_L and P_U in PB confidence intervals are revised as

$$P_L = \Phi\left(z_0 + \frac{z_0 + z_\alpha}{1 - a(z_0 + z_\alpha)} \right) \tag{9}$$

$$P_U = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha}}{1 - a(z_0 + z_{1-\alpha})} \right) \tag{10}$$

$$a = \sum_{i=1}^{B} (\hat{\theta}^* - \hat{\theta}^*(i))^3 / 6 \left[\sum_{i=1}^{B} (\hat{\theta}^* - \hat{\theta}^*(i))^2 \right]^{\frac{3}{2}}, \tag{11}$$

$$\hat{\theta}^* = \sum_{i=1}^{B} \hat{\theta}^*_i / B \tag{12}$$

$$z_0 = \Phi^{-1}(\#\{\hat{\theta}^*(b) < \hat{\theta}\}/B) \tag{13}$$

where $\Phi(.)$ is the cumulative standard normal distribution function and z_α denotes the 100 $(1 - 2\alpha)$th percentage point of the standard normal distribution. z_0 and a are labeled the bias-correction and acceleration constants, respectively. Thus the 100 $(1 - 2\alpha)$% BCa confidence interval for θ is obtained as $[\hat{\theta}^*(P_L B), \hat{\theta}^*(P_U B)]$. When z_0 and a equal zero, then the BCa method is the same as the percentile interval method (Efron and Tibshirani, 1993).

The SB confidence interval is easy to calculate but requires normality assumption on the bootstrap distribution (Efron, 1986). If $\hat{\theta}$ is not approximately normal, then the PB confidence interval is preferable. However, bootstrap distributions obtained using only a sample of the complete bootstrap distribution may shift higher or lower than expected (Efron, 1982). The BCPB method was proposed to correct this potential bias (see Efron, 1982 for a complete justification of the method). Furthermore, an improved version of the
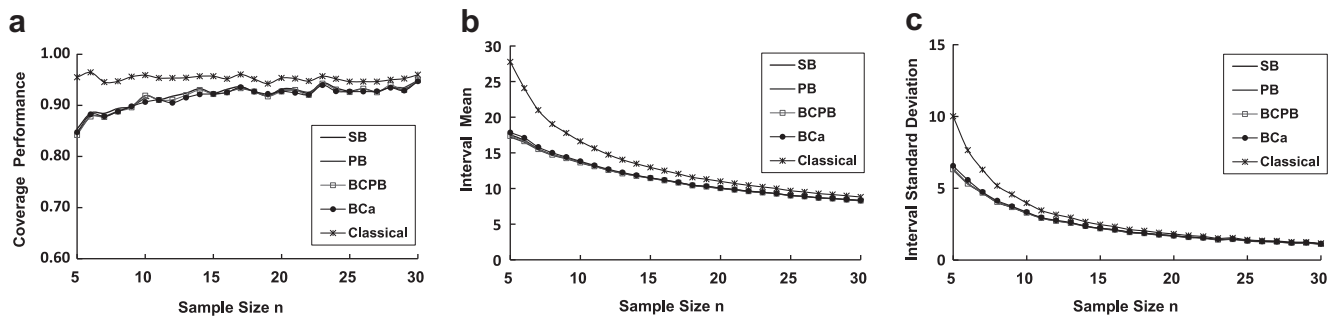
Fig. 2. Coverage performance, interval mean and interval standard deviation vs. sample size under normal distribution with $(\mu, \sigma) = (70, 12)$.

percentile method called BCa (Efron, 1987) can accelerate the correction of the estimated error of biased data in the PB confidence interval. Though no gold standard exists for conclusively identifying which method has the best confidence interval (Zheng, 2002), Efron and Tibshirani (1993) recommends the BCa confidence interval for general use, and this study uses this interval for analysis, particularly for non-parametric problems.

3. Procedure for estimating uncertainty in GHG emission estimates using bootstrap confidence intervals

IPCC (2000, 2006) guidelines suggest using the 95% confidence intervals to express the uncertainty of emission estimates. IPCC (2006) mentioned that in some cases confidence intervals may be positively skewed because of small sample sizes, skewness of the underlying population distribution, or both. IPCC (2006) also suggested that numerical methods, such as bootstrap simulation, can be used instead to obtain the confidence interval in cases where the uncertainty in the mean is not a symmetric distribution. An application of bootstrap approaches is to establish good confidence intervals for estimating uncertainty in GHG inventory. Efron and Tibshirani (1986) indicated that "good" means that the bootstrap confidence intervals have relatively accurate coverage performances and short average interval lengths in all situations (Chu and Ke, 2006). This section presents a procedure for constructing BCa bootstrap confidence intervals to estimate the uncertainty of GHG emission estimates using simulated data from normal, log-normal, and uniform distributions. Furthermore, the procedure for analyzing the accuracy and sensitivity of the bootstrap confidence interval is explained for various sample sizes, mean and standard deviations.

3.1. BCa confidence intervals for estimating uncertainty in GHG emission estimates

The above distributions are commonly used to represent variability in probabilistic assessment regarding estimation of emission

estimate uncertainty (IPCC, 2006). This section sets the distribution to represent inter-source variation in emissions, such that normal distribution denoted by $N(\mu, \sigma)$ of emission estimates is used to generate population data, where μ is the mean of the estimated emission and σ is the standard deviation of the estimated emission. The simulation procedure for BCa confidence interval using bootstrap method is shown in Fig. 1

3.2. Analyzing the validity and sensitivity of bootstrap confidence intervals

The distributions of the inter-source variability emission estimates were set to be normal, log-normal and uniform, respectively, to generate emission data. Accuracy defines the ability to measure the true value of the characteristic correctly on average, and precision refers to the variability in the measurements (Montgomery et al., 2011). The performances of the various bootstrap confidence intervals were evaluated using the following three indices (Chou et al., 2006; Ke et al., 2008; Tong et al., 2008):

1. Coverage performance index: The index represents the percentage of times that the actual emission falls into the bootstrap confidence intervals. For example, if 960 confidence intervals include the actual emission in 1000 bootstrap confidence intervals, then the value of the coverage performance index is 0.96. Larger performance index value indicates more accurate bootstrap confidence interval estimates.
2. Interval mean index: The interval length denotes the difference between the lower and upper limits of the confidence interval. The interval mean index represents an average length of N bootstrap confidence intervals. Moreover, smaller value of interval mean index implies more precise estimation and better performance of the bootstrap confidence interval estimates. The appearance of smaller interval means is associated with smaller coverage performance, therefore it is implied a trade-off relationship between precision and accuracy.
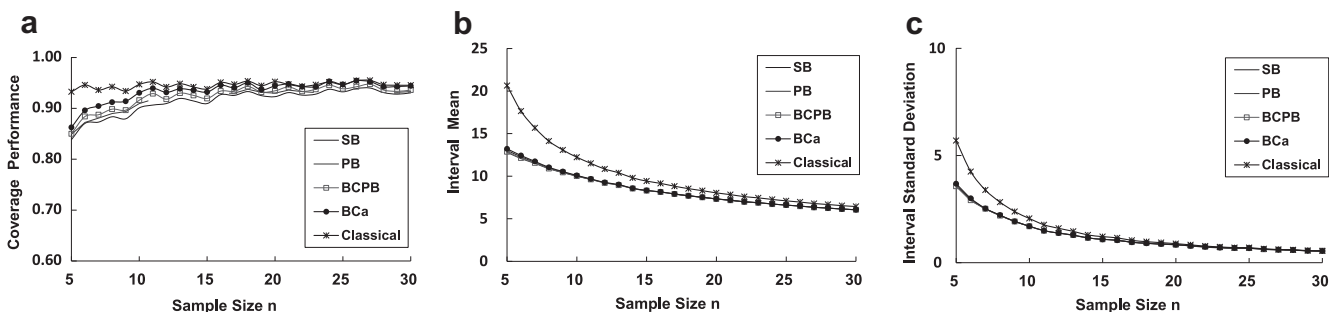

Fig. 3. Coverage performance, interval mean and interval standard deviation vs. sample size under uniform distribution with (Lower Bound, Upper Bound) = (75, 45).
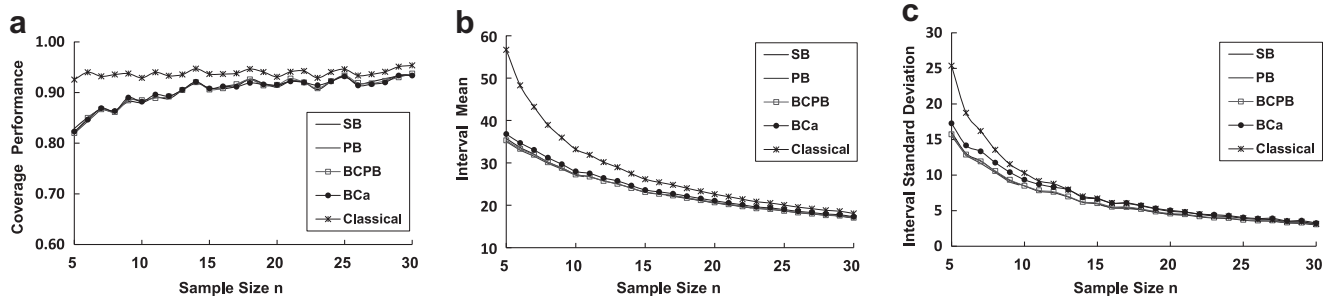
Fig. 4. Coverage performance, interval mean and interval standard deviation vs. sample size under log-normal distribution with $(\mu, \sigma) = (70, 25)$.

3. Interval standard deviation index: The index represents the standard deviation of interval lengths of N bootstrap confidence intervals. A smaller standard deviation implies smaller estimated variation and better performance bootstrap confidence interval estimates.

The procedure for analyzing the accuracy of bootstrap confidence intervals involves repeating steps 2 to 6 in Fig. 1, N times (effectiveness increases with N, and here $N = 2000$) to obtain N sets of bootstrap confidence intervals. The bootstrap intervals can be verified using the coverage performance index.

The procedure for analyzing the sensitivity of bootstrap confidence intervals involves repeating steps 2 to 6 in Fig. 1, N times for various parameter combinations of sample size, mean and standard deviation. The sensitivity of the bootstrap confidence intervals can then be analyzed using the coverage performance, interval mean, and interval standard deviation. The simulation procedure and results are presented as follows.

Monte Carlo simulation was used to generate normally distributed emissions samples with various combinations of parameters, which are listed in Table 1. Similarly, emissions estimates were generated under log-normal and uniform distributions. Table 2 lists the combinations of upper and lower bounds of the uniform distribution. Table 3 lists combinations of emission parameters for log-normal distributions.

For a normal distribution, each pair of parameters (μ, σ), where μ and σ represent the population mean and standard deviation of inter-source variability, respectively, a single sample of size n ($n = 5$, 6, ..., 30) is first randomly taken from the generated emission data. Then, $B = 1000$ bootstrap resamples (each is of size n) are generated from that single sample and the four bootstrap confidence intervals are computed. The single simulation run is then replicated $N = 2000$ times. The three indexes aforementioned are obtained from 2000 bootstrap confidence intervals for evaluating the accuracy and sensitivity of the four bootstrap confidence intervals and the 95% confidence interval of emission estimates. All simulation runs are performed using Excel VBA and Matlab.

Figs. 2–4 plot the three indices versus sample size n for three distributions, namely normal, log-normal and uniform. Classical, SB, PB, BCPB and BCa in the figures represent classical confidence interval, standard bootstrap, percentile bootstrap, biased-corrected percentile bootstrap and bias-corrected accelerated, respectively. Note that, the scales for the vertical min and max of the interval mean and standard deviation in Figs. 2–4 differ because parameters are different for the three distributions. The coverage performance improves with increasing sample size n ($n = 5, 6, ..., 30$), while both the interval mean and standard deviation decrease. Increased sample size thus improves estimation precision and accuracy for the four bootstrap confidence intervals and the classical confidence interval. Figs. 2–4 indicate that the coverage performance is approximately 0.90 or above for sample size n exceeding nine. Even though all four methods present a similar trend in terms of coverage performance, the SB method slightly outperforms the PB, BCPB and BCa methods in terms of normally and log-normally distributed emission estimates. Additionally, the BCa and BCPB methods marginally outperform the SB and PB methods in terms of uniformly distributed emission estimates. The small differences in results when comparing these methods are of little practical significance in the cases explored here. The classical confidence intervals are close to 95% coverage with various sample sizes for the normal population case. But for non-normal cases, the coverage performances of classical confidence interval are lower than 95% when sample size is small.

The four bootstrap approaches are scrutinized using interval mean and interval standard deviation. Figs. 2–4b show that all interval means for the four bootstrap approaches decrease with sample size n. The differences among the average lengths of the four bootstrap confidence intervals are negligible. Figs. 2–4c show that all interval standard deviations for the four bootstrap approaches decrease with sample size n. The differences among the interval standard deviations of the four bootstrap confidence intervals are negligible. Therefore, this study concludes that all the four bootstrap confidence intervals can be applied to estimate the uncertainty of emission estimates − regardless of the normality of the data distribution.

Table 4
Carbon stock of Japanese cedar in Taiwan.

Cited volume of woody biomass (V) (m³ ha⁻¹)	Basic wood density of the extracted wood (D) (tonnes d.m. m⁻³)	BEF	Woody biomass originates (R)	Carbon fraction of dry matter (CF) (tonnes C (tonne d.m.)⁻¹)	Carbon stock (C) (tonnes yr⁻¹)
170.148	0.36	1.23	0.28	0.4903	47.28
180.810	0.36	1.23	0.28	0.4903	50.25
184.082	0.36	1.23	0.28	0.4903	51.16
203.261	0.36	1.23	0.28	0.4903	56.49
205.278	0.36	1.23	0.28	0.4903	57.05
213.777	0.36	1.23	0.28	0.4903	59.41
242.254	0.36	1.23	0.28	0.4903	67.32
248.953	0.36	1.23	0.28	0.4903	69.18
251.369	0.36	1.23	0.28	0.4903	69.85
252.542	0.36	1.23	0.28	0.4903	70.18

Table 5
Confidence intervals for carbon stock of Japanese cedar (tonnes yr^{-1}).

Method	95% confidence interval		Interval length
	Lower limit	Upper limit	
Classical	53.53	66.10	12.57
SB	55.21	64.87	9.66
PB	55.31	64.65	9.34
BCPB	55.31	64.65	9.34
BCa	55.29	64.61	9.32

4. Case study

The proposed methods for estimating uncertainties using bootstrap confidence intervals are illustrated using a case study involving the carbon stock of Japanese cedar in Taiwan. The case study attempts to estimate bootstrap confidence intervals for the carbon stock of Japanese cedar in Taiwan. The Carbon stock evaluation of a forest can be calculated using a multiplicative equation (IPCC, 2003),

$$C = [V \times D \times BEF] \times (1 + R) \times CF \qquad (14)$$

Where C represents forest carbon stock (tonnes yr^{-1}); V represents the volume of woody biomass (m^3 ha^{-1}); D represents the basic wood density of the extracted wood (tonnes d.m. m^{-3}); BEF represents the biomass expansion factor for converting the biomass of extracted round wood to total above-ground tree biomass, dimensionless; R represents the average root-to-shoot ratio, dimensionless; and CF represents carbon fraction of dry matter (default = 0.5), tonnes C (tonne d.m.)$^{-1}$.

The cited data for V, D, BEF, and CF in Eq. (14) came from the Third Survey of Forest Resources and Land Use in Taiwan (Forestry Bureau, CoA, 1995) and previous studies (Lin et al., 2002; Hsieh et al., 2010). The carbon stock of Japanese cedar shown in Table 4 can be calculated using Eq. (14) and the cited data.

Based on the construction of the BCa confidence interval, steps 3–6 from Fig. 1 were performed without any distribution assumption. The remaining three bootstrap intervals (SB, PB and BCPB) are constructed using a similar procedure. Table 5 lists the calculation results of the classical confidence interval and four bootstrap intervals for carbon stock of Japanese cedar. The interval lengths of the bootstrap intervals are smaller than obtained using the classical method, indicating that the smaller interval lengths have better precise for estimating the emission estimate. Therefore, the uncertainty estimate of carbon stock of Japanese cedar in Taiwan ranges approximately ±4.7 (tonnes yr^{-1}).

5. Conclusions

Classical statistical methods are commonly used for estimating uncertainties of GHG emission estimates. However, since the emission data is usually inadequate or its distribution is unknown or non-normal, classical statistical methods fall short in this respect, causing significant bias in uncertainty estimation.

This study presents the results of a simulation study examining the behavior of four 95% bootstrap confidence intervals (namely SB, PB, BCPB and BCa) together with the classical confidence interval for assessing the uncertainty of emission estimates. A comprehensive simulation for a particular combination of sample size and parameters is run under each of the possible distributions, including normal, log-normal and uniform. The accuracy and sensitivity of the uncertainty for various interval estimations are examined by comparing three indices: coverage performance, interval mean and interval standard deviation. Based on the simulation results, this study concludes that regarding the effect of

sample size, large sample size always results in higher coverage performance, shorter interval mean, and smaller interval standard deviation of the bootstrap confidence interval. Increased sample size improves the estimation precision and accuracy for the four confidence intervals. Furthermore, the bootstrap intervals are more applicable than the 95% confidence interval given non-normal dataset and small sample size. When the sample size n is less than 30, the bootstrap confidence interval has a smaller interval length with a smaller deviation than the classical 95% confidence interval regardless of the normality of the data distribution. A sample size greater than or equal to 9 and the bootstrap confidence interval are recommended for estimating the uncertainty of emission estimates. When sample size n exceeds 30, similar results are obtained using either the 95% confidence interval or bootstrap confidence intervals regardless of whether the data distribution is normal or non-normal (Zheng and Frey, 2004). In practice, the emission data are difficult to obtain, and their distribution is not easily determined using the goodness-of-fit test. The proposed method thus can reduce the estimate bias from limited emission data. Future studies can focus on smaller samples of less than nine.

References

Chou, C.-Y., Lin, Y.-C., et al., 2006. On the bootstrap confidence intervals of the process incapability index Cpp. Reliability Engineering & System Safety 91 (4), 452–459.
Chu, Y.K., Ke, J.C., 2006. Confidence intervals of mean response time for an M/G/1 queueing system: bootstrap simulation. Applied Mathematics and Computation 180 (1), 255–263.
Efron, B., Gong, G., 1983. A leisurely look at the bootstrap, the Jackknife, and cross-validation. American Statistician 37, 36–48.
Efron, B., Tibshirani, R., 1986. Bootstrap method for standard efforts, confidence intervals, and other measures of statistical accuracy. Statistical Science 1, 54–77.
Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Champman & Hall, London, U.K.
Efron, B., 1979. Bootstrap method: another look at the Jackknife. The Analysis of Statistics 7 (1), 1–26.
Efron, B., 1982. The Jackknife, The Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics CBMS-NSF Monographs 38, Philadelphia, PA USA.
Efron, B., 1986. How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association 81, 461–470.
Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82, 171–200.
Forestry Bureau, Council of Agriculture executive Yuan, 1995. Third Survey of Forest Resources and Land Use in Taiwan.
Frey, H.C., Bammi, S., 2002. Quantification of variability and uncertainty in lawn and garden equipment NO$_x$ and total hydrocarbon emission factors. Journal of the Air & Waste Management Association 52 (4), 435–448.
Frey, H.C., Li, S., 2003. Quantification of variability and uncertainty in AP-42 emission factors: case studies for natural gas-fueled engines. Journal of the Air & Waste Management Association 53 (12), 1436–1447.
Frey, H.C., Rhodes, D.S., 1996. Characterizing, simulating, and analyzing variability and uncertainty: an illustration of methods using an air toxics emissions example. Human and Ecological Risk Assessment 2, 762–797.
Frey, H.C., Zhao, Y., 2004. Quantification of variability and uncertainty for air toxic emission inventories with censored emission factor data. Environmental Science and Technology 38 (22), 6094–6100.
Frey, H.C., Zheng, J., 2002. Method for Development of Probabilistic Emission Inventories: Example Case Study for Utility NO$_x$ Emissions, Proceedings. U.S. Environmental Protection Agency Emission Inventory Conference, Atlanta, GA.
Gunter, B., 1991. Bootstrapping: how to make something from almost nothing and get statistically valid answers. Part I. Quality Progress, 97–103.
Gunter, B., 1992. Bootstrapping: how to make something from almost nothing and get statistically valid answers. Part III. Quality Progress, 119–122.
Hsieh, H.-C., Wang, D.-H., et al., 2010. Applying FORMOSAT II Image and Field Plots Survey Data to Estimating Carbon Sequestration in Fraxinus Farm Afforestation. Proceedings of the Asia GIS 2010 International Conference, Kaohsiung, Taiwan.
IPCC, 2000. Good practice guidance and uncertainty management in National Greenhouse Gas Inventories. In: Penman, J., Kruger, D., Galbally, I., Hiraishi, T.,

Nyenzi, B., Emmanuel, S., Buendia, L., Hoppaus, R., Martinsen, T., Meijer, J., Miwa, K., Tanabe, K. (Eds.), Intergovernmental Panel on Climate Change IPCC. IPCC/OECD/IEA/IGES, Hayama, Japan.

IPCC, 2003. Good practice guidance for land use, land-use change and forestry. In: Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F. (Eds.), Intergovernmental Panel on Climate Change IPCC. IPCC/IGES, Hayama, Japan.

IPCC, 2006. IPCC guidelines for National greenhouse gas inventories. In: Eggleston, H.S., Buendia, L, Miwa, K., Ngara, T., Tanabe, K. (Eds.), Intergovernmental Panel on Climate Change IPCC. IPCC/OECD/IEA, Hayama, Japan.

Ke, J.-C., Chu, Y.-K., et al., 2008. Comparing confidence limits for short-run process incapability index Cpp. Physica A: Statistical Mechanics and its Applications 387 (13), 3227–3238.

Kini, M.D., Frey, H.C., 1997. Probabilistic Evaluation of Mobile Source Air Pollution. In: Probabilistic Modeling of Exhaust Emission from Light Duty Gasoline Vehicles, vol. 1. Prepared by North Carolina State University for Center for Transportation and the Environment, Raleigh.

Lin, Y.J., Liu, C.P., Lin, J.C., 2002. Measurement of specific gravity and carbon content of important timber species in Taiwan. Taiwan Journal for Science 17 (3), 291–299.

Meeden, G., 1999. Interval estimators for the population mean for skewed distributions with a small sample size. Journal of Applied Statistics 26 (1), 81–96.

Mendenhall, W., Sincich, T., 2007. Statistics for Engineering and the Sciences. Pearson Prentice-Hall, Upper Saddle River.

Montgomery, D.C., Runger, G.C., Hubele, N.F., 2011. Engineering Statistics. Wiley, Hoboken, NJ.

Mooney, C.Z., Duval, R.D., 1993. Bootstrapping: A Nonparametric Approach to Statistical Inference. SAGE, Newbury Park.

Rhodes, D.S., Frey, H.C., 1997. Quantification of Variability and Uncertainty in AP-42 Emission Factors Using Bootstrap Simulation Emission Inventory: Planning for the Future, Air and Waste Management Association Pittsburgh, Pennsylvania, 147–161 pp.

Tong, L.-I., Chen, H.-T., et al., 2008. Constructing BCa bootstrap confidence interval for the difference between two non-normal process capability indices CNpmk. Quality Engineering 20 (2), 209–220.

Young, G.A., 1994. Bootstrap: more than a stab in the dark? Statistical Science 9, 382–415.

Zhao, Y., Frey, H.C., 2004a. Quantification of variability and uncertainty for censored datasets and application to air toxic emission factors. Risk Analysis 24 (3), 1019–1034.

Zhao, Y., Frey, H.C., 2004b. Development of probabilistic emission inventory for air toxic emissions for Jacksonville, Florida. Journal of the Air & Waste Management Association 54 (11), 1405–1421.

Zhao, Y., Frey, H.C., 2006. Uncertainty for data with non-Detects: air toxic emissions from combustion. Human and Ecological Risk Assessment 12 (6), 1171–1191.

Zheng, J., Frey, H.C., 2004. Quantification of variability and uncertainty using Mixture distributions: evaluation of sample size, mixing weights and separation between components. Risk Analysis 24 (3), 553–571.

Zheng, J., 2002. Quantification of variability and uncertainty in emission estimation: general methodology and software implementation. Doctoral dissertation, North Carolina State University, USA.