

Research Article

A Chain Ratio Exponential-Type Compromised Imputation for Mean Estimation: Case Study on Ozone Pollution in Saraburi, Thailand

Kanisa Chodjuntug and Nuanpan Lawson 

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Correspondence should be addressed to Nuanpan Lawson; nuanpan.n@sci.kmutnb.ac.th

Received 23 September 2020; Accepted 16 November 2020; Published 5 December 2020

Academic Editor: Aera Thavaneswaran

Copyright © 2020 Kanisa Chodjuntug and Nuanpan Lawson. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to its impact on health and quality of life, Thailand's ozone pollution has become a major concern among public health investigators. Saraburi Province is one of the areas with high air pollution levels in Thailand as it is an important industrialized area in the country. Unfortunately, the August 2018 Pollution Control Department (PCD) report contained some missing values of the ozone concentrations in Saraburi Province. Missing data can significantly affect the data analysis process. We need to deal with missing data in a proper way before analysis using standard statistical techniques. In the presence of missing data, we focus on estimating ozone mean using an improved compromised imputation method that utilizes chain ratio exponential technique. Expressions for bias and mean square error (MSE) of an estimator obtained from the proposed imputation method are derived by Taylor series method. Theoretical finding is studied to compare the performance of the proposed estimator with existing estimators on the basis of MSE's estimators. In this case study, the results in terms of the percent relative efficiencies indicate that the proposed estimator is the best under certain conditions, and it is then applied to the ozone mean estimation for Saraburi Province in August 2018.

1. Introduction

Air pollution is a global problem which results in negative effects on both the environment and human health. Many researchers have found that air pollution is associated with mortality and morbidity from lung cancer, respiratory, cardiovascular diseases, and exacerbation of chronic respiratory conditions [1, 2]. Moulton and Yang [3] have shown that air pollution is correlated with Alzheimer's disease and other neurodegenerative disorders. From the World Health Organization's report in 2018, air pollution caused approximately 4.2 million deaths [4]. In fact, children are more vulnerable than adults because their lungs, heart, and brain are still growing. Therefore, air pollution is a major public health concern; monitoring and measuring the quality of air is critical.

The Pollution Control Department (PCD) in Thailand is an agency that measures the amount of air pollutants such as sulphur dioxide (SO_2), carbon monoxide (CO), nitrogen oxides (NO_x), particulate matter ($\text{PM}_{2.5}$, PM_{10}), and ozone (O_3). As shown in the PCD's Air Quality Management Division report in 2015, O_3 and PM_{10} concentrations were higher than standard levels in almost every province [5]. There is also a relationship between O_3 and PM_{10} [6, 7]. Naphralan subdistrict, Chaloe Phra Kiat district is an industrialized area in Saraburi Province with traffic congestion and several stone and cement factories. According to the information from the Saraburi-based ground monitoring stations of PCD in August 2018 [8], we found that the O_3 concentration data were missing due to equipment malfunction or errors in measurement. Missing values may cause a significant effect of data analysis process. We deal

with it using standard statistical techniques. In environmental research, a number of techniques can be employed to impute missing values in air pollutants concentration data such as the mean top bottom method, mean imputation method, the multiple regression method, and artificial neural network models [9–13].

In sample surveys, missing values or nonresponses often occur. There are two types of nonresponse: item nonresponse and unit nonresponse. The imputation method is used to handle item nonresponse, and the weight method is applied to deal with unit nonresponse. In addition, imputation which uses available data as a source for replacement of missing data is the most common method to solve missing data.

In addition, many researchers have studied the auxiliary information in order to improve the precision of population mean estimation under a simple random sampling without replacement (SRSWOR). For example, Cochran [14] applied the auxiliary information at the estimation stage and proposed an estimator to estimate the population mean. Bahl and Tuteja [15] first proposed new ratio-type exponential method for estimating the mean of population using information on auxiliary variable, and their methods are more efficient than common methods: mean and ratio methods. Later, Singh and Pal [16] proposed a chain ratio-ratio-type exponential technique which is more efficient than the common estimators including mean, ratio, and ratio-type exponential estimators under certain condition as follows:

$$\bar{y}_{CR} = \bar{y}_n \frac{\bar{X}}{\bar{x}_n} \exp\left(\frac{\bar{X} - \bar{x}_n}{\bar{X} + \bar{x}_n}\right), \quad (1)$$

where \bar{y}_n is sample mean of interest variable Y and \bar{X} and \bar{x}_n are population mean and sample mean of auxiliary variable X , respectively.

Similarly, Lee et al. [17] applied the auxiliary information for the purpose of imputation. Recently, Singh and Horn [18] suggested a compromised imputation method to estimate the population mean as follows:

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1 - \alpha) \hat{b} x_i; & i \in R, \\ (1 - \alpha) \hat{b} x_i; & i \in R^c, \end{cases} \quad (2)$$

where $\hat{b} = (\sum_{i \in R} y_i / \sum_{i \in R} x_i)$, y_i and x_i are observed values of Y and X for the i^{th} unit, R and R^c are the set of responding and nonresponding units, respectively, n and r are sizes of sample and response data, and α is a chosen constant.

Under this imputation method, the point estimator of the population mean becomes

$$\bar{y}_{COMP} = \alpha \bar{y}_r + (1 - \alpha) \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}, \quad (3)$$

where \bar{y}_r , \bar{x}_n , and \bar{x}_r are response mean of the variable of interest Y , sample mean of auxiliary variable X , and response mean of auxiliary variable X , respectively.

The bias and mean square error of \bar{y}_{COMP} are, respectively, given by

$$\text{Bias}(\bar{y}_{COMP}) = (1 - \alpha) \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y} C_X^2 (1 - K), \quad (4)$$

$$\text{MSE}(\bar{y}_{COMP}) = \left(\frac{1}{r} - \frac{1}{N}\right) \bar{Y}^2 C_Y^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y}^2 C_X^2 [(1 - \alpha)^2 - 2(1 - \alpha)K], \quad (5)$$

where $K = \rho(C_Y/C_X)$, C_Y , and C_X are population coefficient of variations of Y and X , respectively. ρ is population correlation coefficient between Y and X . N , n , and r are sizes of population, sample, and response data, respectively. Their research showed that the suitably chosen constant is correlated with the performance of the estimator.

In this study, we aim to use the imputation method to estimate the population mean in the presence of nonresponse occurring in the variable of interest Y only. We propose to improve the compromised imputation method by using the chain ratio-type exponential technique and its corresponding estimator. The bias and the mean square error have been obtained to the first degree of approximation using the Taylor series method. The efficiency of the proposed estimator is compared with some existing estimators on the basis of MSE in order to obtain the certain conditions for application of proposed estimator. In this case study, we use the percent relative efficiency (PRE) as an indicator to

assess the performance of the estimator. Then, the best estimator is applied to estimate the population mean of O_3 as a variable of interest Y based on PM_{10} concentration data as a variable of auxiliary X of the Saraburi Province's data in August 2018.

2. Materials and Methods

2.1. Basic Setup Framework. Let $U = \{1, 2, \dots, N\}$ be a finite population of size N , y_i be the value of interest variable Y , and x_i be value of auxiliary variable X . Let $\bar{Y} = ((\sum_{i=1}^N y_i)/N)$ and $\bar{X} = ((\sum_{i=1}^N x_i)/N)$ be the population means of Y and X , respectively, and they are unknown values. Let R and R^c be the set of responding units and nonresponding units, respectively. The value of y_i is observed for every $i \in R$; meanwhile, the value of y_i is missing for every $i \in R^c$ and imputed value with x_i . Based on SRSWOR scheme, s of size n with paired variable (X, Y) is selected from U and contains

both r responding units and $(n - r)$ nonresponding units. Let $\bar{x}_n = ((\sum_{i=1}^n x_i)/n)$, $\bar{x}_r = ((\sum_{i=1}^r x_i)/r)$ and $\bar{y}_r = ((\sum_{i=1}^r y_i)/r)$ be the sample mean of X and the response mean of X and Y , respectively.

2.2. *Existing Imputation Methods and Corresponding Estimation.* The mean imputation method, the data after imputation, is defined as

$$y_{.i} = \begin{cases} y_i; & i \in R, \\ \bar{y}_r; & i \in R^c, \end{cases} \quad (6)$$

where $\bar{y}_r = ((\sum_{i=1}^r y_i)/r)$.

Under this imputation method, the point estimator of the population mean becomes

$$\bar{y}_s = \frac{\sum_{i \in S} y_{.i}}{r} = \bar{y}_r. \quad (7)$$

The bias and variance of \bar{y}_s are, respectively, given by

$$\text{bias}(\bar{y}_s) = 0, \quad (8)$$

$$\text{var}(\bar{y}_s) = \left(\frac{1}{r} - \frac{1}{N}\right) \bar{Y}^2 C_Y^2, \quad (9)$$

where $C_Y = (S_Y/\bar{Y})$ and $S_Y^2 = ((\sum_{i=1}^N (y_i - \bar{Y})^2)/N - 1)$.

The ratio imputation method, the data after imputation is defined as

$$y_{.i} = \begin{cases} y_i; & i \in R, \\ \hat{b}x_i; & i \in R^c, \end{cases} \quad (10)$$

where $\hat{b} = ((\sum_{i \in R} y_i)/(\sum_{i \in R} x_i))$,

Under this imputation method, the point estimator of the population mean becomes

$$\bar{y}_{\text{RAT}} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}, \quad (11)$$

where $\bar{x}_n = ((\sum_{i=1}^n x_i)/n)$, $\bar{x}_r = ((\sum_{i=1}^r x_i)/r)$, and $\bar{y}_r = ((\sum_{i=1}^r y_i)/r)$ be the sample mean of X and the response mean of X and Y , respectively.

The bias and mean square error of \bar{y}_{RAT} are, respectively, given by

$$\text{bias}(\bar{y}_{\text{RAT}}) = \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y} C_X^2 (1 - K), \quad (12)$$

$$\text{MSE}(\bar{y}_{\text{RAT}}) = \left(\frac{1}{r} - \frac{1}{N}\right) \bar{Y}^2 C_Y^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y}^2 C_X^2 (1 - 2K), \quad (13)$$

where $S_Y^2 = ((\sum_{i=1}^N (y_i - \bar{Y})^2)/N - 1)$, $S_X^2 = ((\sum_{i=1}^N (x_i - \bar{X})^2)/N - 1)$, $S_{XY} = ((\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}))/N - 1)$, $C_Y = (S_Y/\bar{Y})$, $C_X = (S_X/\bar{X})$, $K = \rho(C_Y/C_X)$, and $\rho = (S_{XY}/S_X S_Y)$.

2.3. *Proposed Compromised Imputation Method and Corresponding Estimator.* Motivated by Singh and Pal [16] and Singh and Horn [18], we propose a new compromised imputation method by using the idea of chain ratio-type exponential estimator. The data after imputation are defined as

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1 - \alpha) \frac{x_i}{\bar{x}_r} \bar{y}_r \exp\left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right); & i \in R, \\ (1 - \alpha) \frac{x_i}{\bar{x}_r} \bar{y}_r \exp\left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right); & i \in R^c, \end{cases} \quad (14)$$

where α denotes suitably chosen constants in order that MSE of the estimator is minimum. $\bar{x}_n = ((\sum_{i=1}^n x_i)/n)$, $\bar{x}_r = ((\sum_{i=1}^r x_i)/r)$, and $\bar{y}_r = ((\sum_{i=1}^r y_i)/r)$ be the sample mean of X and the response mean of X and Y , respectively.

Under the proposed imputation method, the point estimator of the population mean is given as follows:

$$\bar{y}_{\text{PE}} = \alpha \bar{y}_r + (1 - \alpha) \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \exp\left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right). \quad (15)$$

Note that if $\alpha = 1$, then $\bar{y}_{\text{PE}} = \bar{y}_r$, and if $\alpha = 0$, then the \bar{y}_{PE} is analogue of the estimator for population mean proposed by Singh and Pal [16].

To find the properties of the proposed estimator, both bias and MSE of \bar{y}_{WTP} are considered up to the first degree of approximation by using the Taylor series method. We define $\bar{y}_r = \bar{Y}(1 + e_1)$, $\bar{x}_n = \bar{X}(1 + e_2)$, $\bar{x}_r = \bar{X}(1 + e_3)$, $M_1 = (1/r) - (1/N)$, $M_2 = (1/n) - (1/N)$, and $M_3 = M_1 - M_2 = (1/r) - (1/n)$, $M_i > 0$; $i = 1, 2, 3$.

Since, SRSWOR is being followed, and we have $E(e_i) = 0$, $i = 1, 2, 3$,

$$\begin{aligned} E(e_1^2) &= M_1 C_Y^2, \\ E(e_2^2) &= M_2 C_X^2, \\ E(e_3^2) &= M_1 C_X^2, \end{aligned} \quad (16)$$

$$E(e_1 e_2) = M_2 \rho C_X C_Y,$$

$$E(e_2 e_3) = M_2 C_X^2,$$

$$E(e_1 e_3) = M_1 \rho C_X C_Y,$$

where $S_Y^2 = ((\sum_{i=1}^N (y_i - \bar{Y})^2)/N - 1)$, $S_X^2 = ((\sum_{i=1}^N (x_i - \bar{X})^2)/N - 1)$, $S_{XY} = ((\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}))/N - 1)$, $\rho = (S_{XY}/S_X S_Y)$, $C_Y = (S_Y/\bar{Y})$, and $C_X = (S_X/\bar{X})$.

Naturally, we can use $s_y^2 = ((\sum_{i=1}^r (y_i - \bar{y}_r)^2)/r - 1)$, $s_x^2 = ((\sum_{i=1}^r (x_i - \bar{x}_r)^2)/r - 1)$, $s_{xy} = ((\sum_{i=1}^r (x_i - \bar{x}_r)(y_i - \bar{y}_r))/r - 1)$, $c_y = (s_y/\bar{y}_r)$, $c_x = (s_x/\bar{x}_r)$, and $\hat{\rho} = (s_{xy}/s_x s_y)$ to estimate S_Y^2 , S_X^2 , S_{XY} , C_Y , C_X , and ρ as population parameters when we cannot find these population parameters.

Next, writing \bar{y}_{PE} in term of e_i 's, equation (15) takes the form as follows:

$$\bar{y}_{PE} = \alpha \bar{Y}(1 + e_1) + (1 - \alpha) \bar{Y}(1 + e_1) \frac{\bar{X}(1 + e_2)}{\bar{X}(1 + e_3)} \exp\left(\frac{\bar{X}(1 + e_2) - \bar{X}(1 + e_3)}{\bar{X}(1 + e_2) + \bar{X}(1 + e_3)}\right). \tag{17}$$

From equation (17), we have

$$\begin{aligned} \bar{y}_{PE} - \bar{Y} &= \alpha \bar{Y}(1 + e_1) + (1 - \alpha) \bar{Y}(1 + e_1) \frac{\bar{X}(1 + e_2)}{\bar{X}(1 + e_3)} \exp\left(\frac{\bar{X}(1 + e_2) - \bar{X}(1 + e_3)}{\bar{X}(1 + e_2) + \bar{X}(1 + e_3)}\right) - \bar{Y} \\ &\approx \alpha \bar{Y}(1 + e_1) + (1 - \alpha) \bar{Y} \left(1 + e_1 + \frac{3}{2} e_2 + \dots + \frac{15}{8} e_3^2\right). \end{aligned} \tag{18}$$

Taking expectation on both sides of equation (18), we get bias (\bar{y}_{PE}) as follows:

$$\text{bias}(\bar{y}_{PE}) = (1 - \alpha) \bar{Y} M_3 \frac{3}{8} (5C_X^2 - 4\rho C_X C_Y). \tag{19}$$

Squaring both sides of equation (18), expanding the term and taking expectations, and retaining the terms up to first degree of approximations, we get MSE (\bar{y}_{PE}) as follows:

$$\text{MSE}(\bar{y}_{PE}) = \bar{Y}^2 \left[M_1 C_Y^2 + 3(1 - \alpha) M_3 C_X \left\{ \frac{3}{4} (1 - \alpha) C_X - \rho C_Y \right\} \right]. \tag{20}$$

2.4. Estimation of Optimum Value Constant (α_{opt}). In this section, we consider α that is optimum in order to find the minimum MSE (\bar{y}_{PE}). Since \bar{y}_{PE} as given in equation (20) is a function of unknown constant α , it needs to search for α optimum values, such that MSE (\bar{y}_{PE}) becomes minimum value. To obtain the constant α , we differentiate equation (20) with respect to α and equate it to zero as follows:

$$\frac{d\text{MSE}(\bar{y}_{PE})}{d\alpha} = \bar{Y}^2 \left[3M_3 C_X \left\{ \rho C_Y - \frac{3}{2} (1 - \alpha) C_X \right\} \right] = 0. \tag{21}$$

We solve equation (21), and we get

$$\alpha_{opt} = 1 - \frac{2}{3} \frac{\rho C_Y}{C_X}. \tag{22}$$

3. Efficiency Comparison of the Proposed Estimator

Under optimum value constant in equation (22), comparison of \bar{y}_{PE} with those estimators including \bar{y}_S , \bar{y}_{RAT} , and \bar{y}_{COMP} is carried out by using MSE and an estimator with a preferred smaller MSE value. We can observe the efficiency of the proposed estimator as follows:

From equations (9) and (20), we have

$$\text{var}(\bar{y}_S) - \text{MSE}(\bar{y}_{PE}) = M_3 \bar{Y}^2 3C_X (1 - \alpha) \left(\rho C_Y - \frac{3}{4} (1 - \alpha) C_X \right) > 0, \tag{23}$$

if $1 - (4/3)\rho(C_Y/C_X) < \alpha < 1$ or $1 < \alpha < 1 - (4/3)\rho(C_Y/C_X)$.

From equations (13) and (20), we have

$$\text{MSE}(\bar{y}_{RAT}) - \text{MSE}(\bar{y}_{PE}) = M_3 \bar{Y}^2 C_X \left(\left[1 - \frac{9}{4} (1 - \alpha)^2 \right] C_X - [2 - 3(1 - \alpha)] \rho C_Y \right) > 0, \tag{24}$$

if $(1/3) < \alpha < 1 - (2/3)(2\rho(C_Y/C_X) - 1)$ or $1 - (2/3)(2\rho(C_Y/C_X) - 1) < \alpha < (1/3)$.

From equations (5) and (20), we have

$$\text{MSE}(\bar{y}_{COMP}) - \text{MSE}(\bar{y}_{PE}) = M_3 \bar{Y}^2 C_X (1 - \alpha) \left(\rho C_Y - \frac{5}{4} (1 - \alpha) C_X \right) > 0, \tag{25}$$

TABLE 1: PRE (\cdot) of each estimator with respect to \bar{y}_S and conditions of α in which \bar{y}_{PE} is better than others.

Estimators	Conditions of α in which \bar{y}_{PE} is better than	PRE (\bar{y})
\bar{y}_S	$1 < \alpha < 2.18$	100.00
\bar{y}_{RAT}	$0.33 < \alpha < 2.84$	84.97
\bar{y}_{COMP}	$1 < \alpha < 1.71$	104.64
\bar{y}_{PE}	$\alpha = \alpha_{opt} = 1.59$	105.25

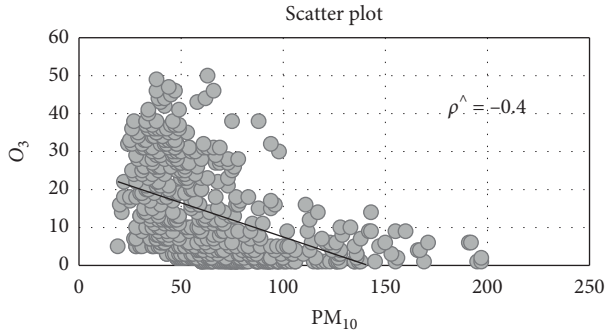


FIGURE 1: The scatter plot of O_3 and PM_{10} data.

if $1 - (4/5)\rho(C_Y/C_X) < \alpha < 1$ or $1 < \alpha < 1 - (4/5)\rho(C_Y/C_X)$.

When the conditions in equations (23)–(25) are satisfied, \bar{y}_{PE} is more efficient than \bar{y}_S , \bar{y}_{RAT} , and \bar{y}_{COMP} , respectively.

4. Case Study

For the case study, we obtained data O_3 level ppb and PM_{10} level ($\mu\text{g}/\text{m}^3$) on a time-scale of one per hour (hourly average) from the PCD website in August 2018. The data belong to the population of 744 units. On examination, we found that the O_3 concentration data had missing values, so it was taken as a variable of interest Y , and PM_{10} concentration data was taken as a variable of auxiliary X . The following values were obtained for the considered variables: $N = 744$, $n = 580$, $r = 545$, $\bar{y}_r = 13.55$, $\bar{x}_n = 69.33$, $\bar{x}_r = 66.29$, $s_y = 12.45$, $s_x = 32.65$, $c_y = 0.92$, and $c_x = 0.49$. We identified 6.03% of the O_3 data as missing.

The conditions of α for which \bar{y}_{PE} is better than the existing estimators are shown in Table 1. The table also presents the percent relative efficiencies (PREs) of each estimator (\bar{y}) with respect to \bar{y}_S which can be computed by $\text{PRE}(\bar{y}) = ((\text{Var}(\bar{y}_S))/(\text{MSE}(\bar{y}))) \times 100$.

The scatter plot (Figure 1) indicates that O_3 and PM_{10} concentration data have a negative relationship. The correlation coefficient between O_3 and PM_{10} concentration data is -0.47 . From Table 1, we consider both the certain conditions of α and PREs; accordingly, \bar{y}_{PE} is more efficient than others. \bar{y}_{PE} is the most suitable for estimating the mean value of O_3 concentration in this case study. The mean value of O_3 concentration by using \bar{y}_{PE} is equal to 13.00 ppb/hr or 0.01 ppm/hr which does not exceed the average O_3 standard of the PDC in Thailand (≤ 0.10 ppm/hr). The mean square error of \bar{y}_{PE} is 0.07 which is close to zero, so it indicates that the proposed estimator is effective in this case.

5. Conclusions

In this case study, when missing data occur in the variable of interest, we propose the improved compromised imputation method using the chain ratio-type exponential technique for the population mean estimation. The mean square error of the proposed estimator was studied under general and optimum situations. We suggest this proposed method is useful for estimating the population mean in the presence of common, real-world, nonresponse data, and it is efficient for applying to the real dataset with missing values under certain condition of α . In fact, in this case study, we applied our proposed method to the whole process of ozone population mean estimation from real data. On the contrary, the common methods proposed by the other authors have not applied their techniques to estimate population mean from real data. In addition, when a dataset contains some missing values, our proposed method can handle this problem to complete data and can also save both time and budget for research conduction. Therefore, the proposed method is a good strategy to apply in practice in case that a missing data problem occurs. However, this study focused on missing values in the variable of interest only, but we could apply a similar method to cases where missing data happens in both the interest and auxiliary variables.

Data Availability

The data to support this study are available on the website of the Pollution Control Department (PCD) of Thailand (<http://air4thai.pcd.go.th/webV2/history/>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the Pollution Control Department (PCD) of Thailand for supporting the data in this research.

References

- [1] A. J. Cohen, H. Ross Anderson, B. Ostro et al., "The global burden of disease due to outdoor air pollution," *Journal of Toxicology and Environmental Health, Part A*, vol. 68, no. 13–14, pp. 1301–1307, 2005.
- [2] Y. Guo, S. Li, B. Tawatsupa, K. Punnasiri, J. J. K. Jaakkola, and G. Williams, "The association between air pollution and mortality in Thailand," *Scientific Reports*, vol. 4, Article ID 5, 2014.
- [3] P. V. Moulton and W. Yang, "Air pollution, oxidative stress, and Alzheimer's disease," *Journal of Environmental and Public Health*, vol. 2012, Article ID 472751, 2012.
- [4] World Health Organization, *Ambient Air Pollution*, World Health Organization, Geneva, Switzerland, 2019, [https://www.who.int/en/newsroom/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/en/newsroom/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health).

- [5] Pollution Control Department of Thailand, *Thailand State of Pollution Report 2015*, http://infofile.pcd.go.th/mgt/PollutionReport2015_en.pdf, Pollution Control Department of Thailand, Bangkok, Thailand, 2018, http://infofile.pcd.go.th/mgt/PollutionReport2015_en.pdf.
- [6] Y. Ding, G. Liu, X. Guo et al., "Correlation analysis between ozone and other atmosphere characteristics," *Chemical Engineering Transactions*, vol. 61, pp. 1747–1752, 2017.
- [7] T. Nishanth, K. M. Praseed, M. K. S. Kumar, and K. T. Valsaraj, "Influence of ozone precursors and PM10 on the variation of surface O₃ over Kannur, India," *Atmospheric Research*, vol. 138, no. 3, pp. 112–124, 2014.
- [8] Pollution Control Department of Thailand, *Thailand's Air Quality and Situation Report*, Pollution Control Department of Thailand, Bangkok, Thailand, 2018, <http://air4thai.pcd.go.th/webV2/history/>.
- [9] S. Ali and S. Dacey, "Technical review: performance of existing imputation methods for missing data in SVM ensemble creation," *International Journal of Data Mining and Knowledge Management Process*, vol. 7, no. 6, pp. 75–91, 2017.
- [10] A. Arroyo, A. Herrero, V. Trico, E. Corchado, and M. Woźniak, "Neural models for imputation of missing ozone data in air-quality dataset," *Complexity*, vol. 2018, Article ID 7238015, , 2018.
- [11] M. D. N. F. Fitri, N. A. Ramli, A. S. Yahaya et al., "Monsoonal differences and probability distribution of PM₁₀ concentration," *Environmental Monitoring and Assessment*, vol. 163, no. 1, pp. 655–667, 2010.
- [12] N. M. Noor, A. S. Yahaya, N. A. Ramli, and M. M. A. B. Abdhullah, "The replacement of missing values of continuous air pollution monitoring data using mean top bottom imputation technique," *Journal of Engineering Research and Education*, vol. 3, pp. 96–105, 2006.
- [13] M. N. Norazian, Y. A. Shukri, R. N. Azam et al., "Estimation of missing values in air pollution data using single imputation technique," *ScienceAsia*, vol. 34, pp. 341–345, 2008.
- [14] W. G. Cochran, *Sampling Technique*, John Wiley and Sons, New York, NY, USA, 3rd edition, 1997.
- [15] S. Bahl and R. K. Tuteja, "Ratio and product type exponential estimators," *Journal of Information and Optimization Sciences*, vol. 12, no. 1, pp. 159–164, 1991.
- [16] H. P. Singh and S. K. Pal, "A new chain ratio-ratio type exponential estimator using auxiliary information in sample surveys," *International Journal of Mathematics and its Applications*, vol. 3, no. 4, pp. 37–46, 2015.
- [17] H. Lee, E. Rancourt, and C. E. Sarndal, "Experiments with variance estimation from survey data with imputed values," *Journal of Official Statistics*, vol. 10, no. 3, pp. 231–243, 1994.
- [18] S. Sing and S. Horn, "Compromised imputation in survey sampling," *Metrika*, vol. 51, pp. 266–276, 2000.