# Modeling generalized statistical distributions of PM2.5 concentrations during the COVID-19 pandemic in Jakarta, Indonesia

*by* Prayudhy Yushananta

---

# Modeling generalized statistical distributions of PM$_{2.5}$ concentrations during the COVID-19 pandemic in Jakarta, Indonesia

**Warsono Warsono[a*], Yeftanus Antonio[b], Slamet B. Yuwono[c], Dian Kurniasari[a], Erdi Suroso[d], Prayudhy Yushananta[e], Mustofa Usman[a] and Sutopo Hadi[f]**

[a]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia
[b]Alumnus of Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia
[c]Department of Forestry, Faculty of Agriculture, Universitas Lampung, Indonesia
[d]Department of Agricultural Industry Technology, Faculty of Agriculture, Universitas Lampung, Indonesia
[e]Department of Environmental Health, Poltekkes Kemenkes Tanjungkarang, Indonesia
[f]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

| CHRONICLE | ABSTRACT |
|---|---|
| | Understanding the probabilistic or statistical behavior of air concentrations is necessary for the effective management of air pollution, such as PM$_{2.5}$. Failure to consider the appropriateness of the model can lead to making inferences that are not supported by scientific evidence. The main focus of this article is to find the best statistical distribution in fitting PM$_{2.5}$ concentrations in the periods of February–June 2018 and February–June 2019 (the periods without COVID-19) and in the period of February–June 2020 (the period with COVID-19) in Jakarta, Indonesia. This article considers making an assessment of the performance of both generalized distributions (e.g., generalized gamma, generalized extreme value, and generalized log-logistic [GLL]) and classical distributions (such as lognormal [LN], gamma, Weibull, log-logistic, and Gumbel) in modeling daily concentrations of PM$_{2.5}$ in the period of February–June 2020, or the period during which the COVID-19 pandemic is present, in Jakarta. For comparison purposes, this study also analyzed PM$_{2.5}$ concentrations in the periods of February–June 2018 and February–June 2019. The comparative evaluation of the models of each period of data uses graphical analyses and goodness-of-fit statistics. The results of applications indicate that the generalized distributions fit the data better than do the classical distributions. Particularly, compared with the classical distributions, including the LN model, the GLL distribution is the most appropriate model in fitting PM$_{2.5}$ concentrations in the periods without and during the period with COVID-19 in Jakarta, Indonesia. |
| | |

## 1. Introduction

The World Health Organization (WHO, 2020) declared COVID-19 a pandemic in early March 2020. The first cases of COVID-19 in Indonesia were reported on March 02, 2020, of which two cases were from Depok, a city located below Jakarta. In the initial phase, the spread was relatively slow until the second week and concentrated in Jakarta and surrounding areas. However, the numbers began to rise rapidly until the end of March 2020. In order to combat the spread of the virus, the President of the Republic of Indonesia (Government of the Republic of Indonesia, 2020) declared public policies of social distancing, named the Large-Scale Social Restriction (LSSR). After receiving an approval letter from the Health Minister, Jakarta state's Governor declared the LSSR on April 10, 2020. This has led to reduced human outdoor activities. Recent literature has reported air quality improvements associated with social distancing measures. For example, Li *et al.* (2020) reported that during the COVID-19 control period, human activities have decreased greatly, and air pollutants like SO$_2$ and PM$_{2.5}$ in Yangtze Delta Region, China, have been reduced significantly. Similar conditions can be found in the work of Kerimray *et al.* (2020), Otmani *et al.* (2020), Zambrano–Monserrate *et al.* (2020) and Cameletti (2020). Air pollution studies conducted to monitor ambient levels and quantify the concentration of various pollutants, such as

particulate matter (PM), entering a given environmental area are greatly concerned about the possible adverse health effects caused by these pollutants. For example, fine particulate matter smaller than 2.5 μm in aerodynamic diameter ($PM_{2.5}$) becomes a major health burden and plays an increasingly negative role in China's social and economic development (Li *et al.*, 2018) and population mortality in an urbanized valley in the American tropics (Aguiar-Gil *et al.*, 2020). Information about the frequency distribution of pollutants is crucial for air pollution management. When the parent statistical distribution of $PM_{2.5}$ concentrations is correctly selected, the specific distribution can be used, for example, to predict mean concentration, the probability of exceeding a critical concentration, and the return period for setting regulatory targets and issuing environmental alerts for public health (Lu and Fang, 2002; Xi *et al.*, 2013; Plocoste *et al.*, 2020). Statistical distributions can also provide a simple description or summary of masses of data.

A significant number of studies suggesting that empirical distributions of pollutant concentration data tend to be lognormally distributed have been published. Lu (2002) presented a critical review of statistical distributions, such as lognormal (LN), Weibull, and type V Pearson, and stated that the LN distribution has been the most appropriate distribution in representing $PM_{10}$ and $PM_{2.5}$ in Sha-Lu, Taiwan. Recently, Xi *et al.* (2013) have reported that instead of the Weibull and gamma distributions, the LN distribution was the best fit distribution for $PM_{10}$ daily concentration in five cities of China. Other examples of the LN distribution for air quality data can be found in the articles of Kalpasanov and Kurchatova (1976), Kushner (1976), Owen and DeRouen (1980), Mage and Ott (1984), Kao and Friedlander (1995), Lu and Fang (2002). However, other parametric distributions can also be used successfully for air pollutant data. On the basis of the sum-of-squares error, Bencala and Seinfeld (1976) showed that Weibull models produce values lower than those of the LN model for five of eight CO data sets. In fitting $PM_{2.5}$ and $PM_{10}$, Karaca *et al.* (2004) concluded that instead of the LN, gamma, and Weibull distributions, the log-logistic (LL) distribution was found to be the most suitable distribution to represent their statistical characters. Moreover, Marani *et al.* (1986) have demonstrated that in terms of log-likelihood (log-L) values, the generalized gamma (GG) distribution, compared with the LN, Weibull, and gamma distributions, was the most suitable statistical model in modeling air pollutant concentrations in the area of Venice, Italy. Applications of the GG distribution to air pollution problems can also be looked at in the study of Lavagnini and Camuffo (1987). Obviously, in fitting air pollutant data, none of the probability models, including the classical LN, has been identified to be superior to others in a general sense. One approach to overcoming this problem could be to use a very general model that includes most of the distributions as limiting distributions or special cases. Among the general models, the generalized log-logistic (GLL) distribution has good potential for fitting environmental pollutant data. The GLL distribution is an extension of the LL distribution. It contains several well-known distributions, such as LN, Weibull, and GG, as special cases (Singh *et al.*,1988). The main focus of this study, therefore, is to assess the performances of the generalized distributions, especially the GLL model, and the classical distributions in fitting $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia.

## 2. Data and methods

### 2.1. Data Source

This study uses the data of daily $PM_{2.5}$ concentrations in the periods of February–June 2018, February–June 2019, and February–June 2020 in Central Jakarta reported by the Indonesian Department of Meteorology, Climatology and Geophysics (https://aqicn.org/city/indonesia/jakarta/us-consulate/central/). We assigned the periods of February–June 2018 and February–June 2019 as the periods without COVID-19 and the period of February–June 2020 as the period with COVID-19.

### 2.2. Methods

As mentioned earlier in the previous section, of primary interest is the fact that, in fitting air pollutant data, none of the probability models, including the classical LN, has been identified to be superior to others in a general sense. It is reasonable, therefore, to use a rich family of generalized distributions that includes several well-known distributions as special cases for fitting air pollutant data. This study proposes the generalized distributions GLL, GG, and generalized extreme value (GEV) to fit $PM_{2.5}$ concentration data in the periods of February–June 2018, February–June 2019, and February–June 2020 in Jakarta, Indonesia. For comparison purposes, the study also proposes the classical distributions LN, gamma, Weibull, LL, and Gumbel. The probability distribution functions (PDFs) of the compared models can be seen in Table 1. For estimating the parameters of the PDF of the distributions, this study uses the method of maximum-likelihood estimation (MLE). Comparative evaluations of the models of data use graphical methods and numerical values of goodness-of-fit statistics. Graphically, this study utilizes the empirical PDF, cumulative distribution function (CDF), and probability–probability (P–P) plot. In terms of goodness-of-fit statistics, this study proposes the following statistics tests: Kolmogorov–Smirnov (K–S), Cramer von Mises (CvM), Anderson–Darling (A–D), and log-L. For computation, this study operates a distribution fitting package in R called "fitdistrplus" to provide functions for estimating parameters of distributions, producing several figure results, and calculating goodness-of-fit statistics (Delignette-Muller and Dutang, 2015). This package requires that a distribution be defined through three functions: the PDF, the CDF, and the quantile function. Except the GLL distribution, all of the compared distributions are available as functions in R. Thus, the initial step is to define the

GLL distribution, which is not yet available. Furthermore, this function estimates the parameters using the MLE method as a default method. Optimization of the likelihood function is done by default using the Nelder–Mead method (Nelder and Mead, 1965), which is a direct search method. To choose the best distribution model, this package also provides goodness-of-fit statistics by performing the "gofstat" function.

**Table 1**
The probability density function of proposed distributions

| Distribution | Probability Density Function |
|---|---|
| Lognormal | $f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}; \; 0 < x, 0 < \sigma$ |
| Gamma | $f(x) = \dfrac{1}{s^a \Gamma(a)} x^{a-1} e^{-\frac{x}{s}}; \; 0 < x, 0 < s, 0 < \alpha$ <br> $\Gamma(\alpha)$ is the gamma function |
| Weibull | $f(x) = \dfrac{a}{b}\left(\dfrac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}; \; 0 < x, 0 < a, 0 < b$ |
| Log-logistic | $f(x) = \dfrac{\gamma \left(\frac{x}{\theta}\right)^\gamma}{\left[1 + \left(\frac{x}{\theta}\right)^\gamma\right]^2}; \; 0 < x, 0 < \theta, 0 < \gamma$ |
| Gumbel | $f(x) = \dfrac{e^{-\left(\frac{x-a}{b}\right)}}{b} e^{-e^{-\left(\frac{x-a}{b}\right)}}; \; x \in R, b > 0$ |
| Generalized gamma | $f(x) = \dfrac{|Q|(Q^{-2})^{Q^{-2}}}{\sigma x \Gamma(Q^{-2})} e^{\left(Q^{-2}(Qw - e^{Qw})\right)}; 0 < x$ <br> where $w = \dfrac{\log(Q^2 \gamma)}{Q}$ and $x = e^{\mu + \sigma w}$ <br> $\Gamma(Q^{-2})$ is the gamma function |
| Generalized extreme value | $f(x) = \left(1 + \dfrac{s(x-a)}{b}\right)^{\left(-\frac{1}{s}\right)-1} e^{-\left(1 + \frac{s(x-a)}{b}\right)^{-1/s}}$ <br> where $0 < \left(1 + \dfrac{s(x-a)}{b}\right)$ and $0 < b$ |
| Generalized log-logistic | $f(x) = \left(\dfrac{\alpha}{x\,B(m_1, m_2)}\right)[G(x)]^{m_1}[1 - G(x)]^{m_2}; 0 < x, 0 < m_1, 0 < m_2$ <br> $B(m_1, m_2)$ is the beta function and $G(x) = \left(1 + e^{-(\beta + \alpha \ln(x))}\right)^{-1}$ |

## 3. Results and discussion

### 3.1. Descriptive Data

As an initial step in assessing the frequency distributions of $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia, this study explores graphically these data by using time series plots, histograms and boxplots. Fig. 1 provides a plot of the daily variation of the time series data of $PM_{2.5}$ concentrations in the periods of February–June 2018, February–June 2019, and February–June 2020 in Jakarta. Fig. 2 and Fig. 3 present, respectively, the histograms and boxplots of $PM_{2.5}$ concentrations in the periods without and during the period with COVID-19.
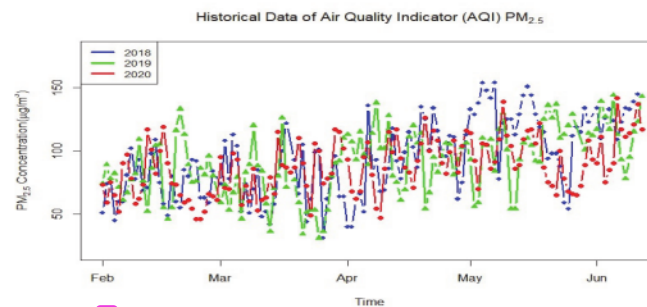


**Fig. 1**. Time series plots of $PM_{2.5}$ concentrations in the periods without and during the period with COVID-19

Fig. 1 presents a random and seasonal pattern of $PM_{2.5}$ concentrations. In particular, $PM_{2.5}$ concentrations increased in April–June of 2018, 2019, and 2020. According to this figure, it can also be seen that several extreme values exist in May 2018 and February 2019. However, such existence was not found in 2020. It should be noted that lower concentrations are seen in 2020 than in 2018 and 2019 at the end of February and in early June. Extreme values in February–June 2020 are less than those in 2018 and 2019 in the same periods as well. On the basis of Figure 2, the frequency distribution of $PM_{2.5}$ concentrations during the period with COVID-19 (February–June 2020) appears somewhat right-skewed and unimodal. But a number of data in the interval right bin of the histogram between 100 and 400 $PM_{2.5}$ concentrations during the period with COVID-19 are less than those in the periods without COVID-19. Moreover, the boxplots in Figure 3 show that the data during the period with COVID-19 are less spread out than those in the period without COVID-19. Values of the third and second quartiles (median) of the data during the period with COVID-19 are slightly less than those in the periods without COVID-19. This is one indication of an improving trend of air quality in Jakarta.
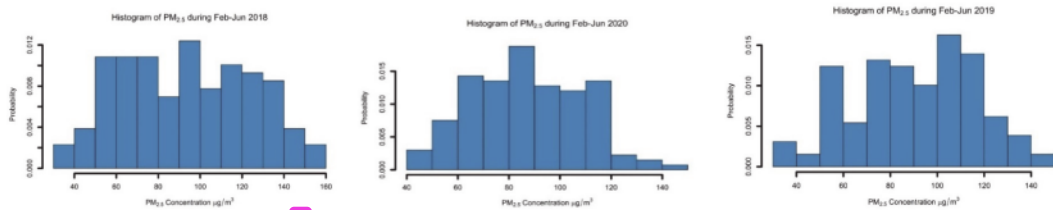


**Fig. 2.** Histograms of $PM_{2.5}$ concentrations in the periods without and during the period with COVID-19
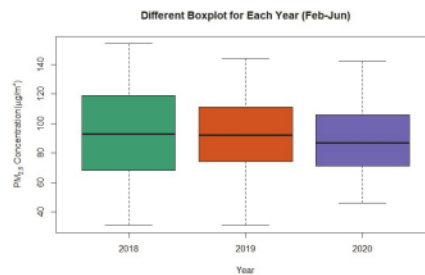


**Fig. 3**. Boxplots of $PM_{2.5}$ concentrations in the periods without and during the period with COVID-19

*3.2. Fitting Distribution*

In order to fit the frequency distributions of PM2.5 concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia, the classical distributions (LN, gamma, Weibull, LL, and Gumbel) and the generalized distributions (GG, GEV, and GLL) were executed. For computing and creating graphs, this work utilized the script of R packages. Figure 4 shows the graphs of the CDFs of the compared models superimposed on the empirical distribution function, which is defined as the proportion of sample observations less than or equal to the values of data. In an attempt to select the most appropriate distributions, the goodness-of-fit statistics of each distribution using K–S, CvM, A–D, and log-L were calculated. Table 2 provides a summary of the goodness-of-fit statistics of the LN, gamma, Weibull, LL, Gumbel, GG, GEV, and GLL distributions. In order to fit the frequency distributions of PM2.5 concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia, the classical distributions (LN, gamma, Weibull, LL, and Gumbel) and the generalized distributions (GG, GEV, and GLL) were executed.
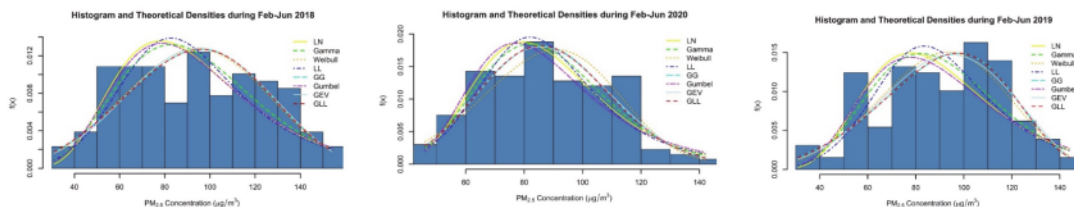


**Fig. 4.** Empirical PDFs of LN, gamma Weibull, LL, Gumbel, GG, GEV, and GLL plot distribution superimposed on relative frequency histograms of $PM_{2.5}$ concentrations in the period of without and during COVID-19

For computing and creating graphs, this work utilized the script of R packages. One of the results of this study is as follows. Fig. 4 displays graphs of the empirical PDFs of compared models superimposed on relative frequency histograms of data. The graphs suggest that in the first period without COVID-19 (February–June 2018) and during the period with COVID-19 (February– June 2020), the GLL and GEV distributions seem close to the frequency distribution of data. Meanwhile, in the second period without COVID-19 (February–June 2019), the GLL and GG models seem close to the data. Hence, graphically, the GLL, GEV, and GG distributions seem to be better distributions than the LN, gamma, Weibull, LL, and Gumbel distributions in modeling $PM_{2.5}$ concentrations in the period without COVID-19 (February–June 2018) and during the period with COVID-19 (February–June 2020). But in the second period without COVID-19 (February–June 2019), the GLL and GG distributions are the most appropriate models in describing data.

**Table 2**
Goodness-of-fit statistics of the compared distributions in fitting data of $PM_{2.5}$ concentrations in the periods of February–June 2018, February–June 2019, and February–June 2020 in Jakarta, Indonesia

| Year | Distribution | Goodness-of-Fit Statistic | | | |
|---|---|---|---|---|---|
| | | K–S | CvM | A–D | Log-L |
| **2018 (February–June)** | LN | 0.09 (7) | 0.23 (8) | 1.45 (7) | --627.29 (6) |
| | Gamma | 0.09 (7) | 0.17 (7) | 1.10 (5) | −624.29 (5) |
| | Weibull | **0.07 (1)** | 0.14 (2) | **0.87 (1)** | −622.03 (3) |
| | LL | 0.08 (3) | 0.20 (6) | 1.50 (8) | −630.62 (8) |
| | Gumbel | 0.08 (3) | 0.14 (2) | 1.31 (6) | −627.53 (7) |
| | GG | 0.08 (3) | 0.14 (2) | **0.87 (1)** | −622.01 (2) |
| | GEV | 0.08 (3) | 0.15 (5) | 0.99 (4) | **−621.73 (1)** |
| | GLL | **0.07 (1)** | **0.13 (1)** | **0.87 (1)** | −622.05(4) |
| **2019 (February–June)** | LN | 0.10 (8) | 0.34 (8) | 2.13 (8) | −613.69 (7) |
| | Gamma | 0.09 (6) | 0.22 (6) | 1.40 (5) | −608.73 (5) |
| | Weibull | 0.06 (4) | 0.07 (4) | 0.47 (4) | −602.80 (4) |
| | LL | 0.08 (5) | 0.19 (5) | 1.69 (6) | −613.64 (6) |
| | Gumbel | 0.09 (6) | 0.31 (7) | 2.02 (7) | −614.16 (8) |
| | **GG** | **0.05 (1)** | **0.05 (1)** | **0.36 (1)** | −602.37 (2) |
| | GEV | **0.05 (1)** | **0.05 (1)** | 0.38 (3) | **−602.13 (1)** |
| | **GLL** | **0.05 (1)** | **0.05 (1)** | **0.36 (1)** | −602.37 (2) |
| **2020 (February–June)** | LN | 0.08 (7) | 0.17 (6) | 1.09 (6) | −599.73 (6) |
| | Gamma | 0.07 (4) | 0.13 (5) | 0.85 (5) | −598.19 (5) |
| | Weibull | 0.07 (4) | 0.10 (1) | 0.72 (4) | −598.57 (4) |
| | LL | **0.06 (1)** | 0.17 (6) | 1.20 (7) | −603.35 (8) |
| | Gumbel | 0.09 (8) | 0.19 (8) | 1.25 (8) | −602.10 (7) |
| | GG | 0.07 (4) | 0.11 (4) | 0.71 (2) | −597.46 (3) |
| | **GEV** | **0.06 (1)** | **0.10 (1)** | **0.68 (1)** | **−597.07 (1)** |
| | **GLL** | **0.06 (1)** | **0.10 (1)** | 0.71 (2) | −597.50 (2) |

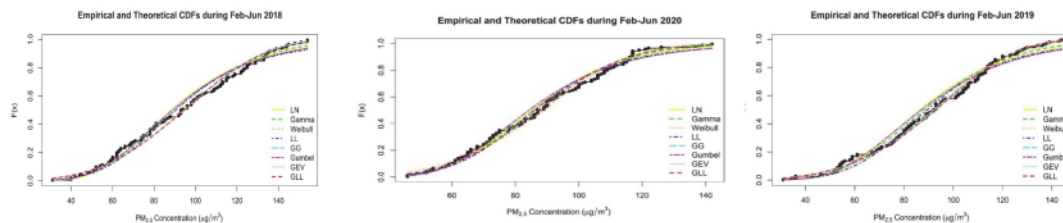Note: Ranking listed in parentheses



**Fig. 5**. The CDF plot of the LN, gamma, Weibull, LL, Gumbel, GG, GEV, and GLL distributions superimposed on the empirical distribution function of $PM_{2.5}$ concentrations in the periods without and during the period with COVID-19

Fig. 5 presents graphs of the CDFs of the compared models superimposed on the empirical distribution function, which is defined as the proportion of sample observations less than or equal to the values of data. From the CDF's standpoint, the graphs suggest that although each compared model performs relatively persistent in fitting $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020), the GLL and GEV distributions performs close to the empirical distribution function of data. Thus, the GLL and GEV distributions fit better than the LN, gamma, Weibull, LL, Gumbel, and GG distributions in the periods without and during the period with the COVID-19 pandemic. Fig. 6 provides the P–P plot of the LN, gamma, Weibull, LL, Gumbel, GG, GEV, and GLL distributions. According to the figure, it is clear that the P–P plot of the GLL distribution is approximately linear, making the GLL distribution the closest model to the linear line in all of the periods of data sets and thus the most suitable model in representing the data of $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020). The smaller the K–S, CvM, and A–D statistics, the better the distribution fits the data. On the other hand, the higher the log-L statistics, the more appropriate the distribution matches the data. According to Table 2, it is clear that the log-L values of the GLL, GG, and

GEV distributions are higher than those of the LN, gamma, Weibull, LL, and Gumbel distributions for all of the three data sets of the periods of February–June 2018, February–June 2019, and February–June 2020. Furthermore, the K–S, CvM, and A–D values of the GLL, GG, GEV distributions are generally smaller than those of the LN, gamma, Weibull, LL, and Gumbel distributions for the three periods.
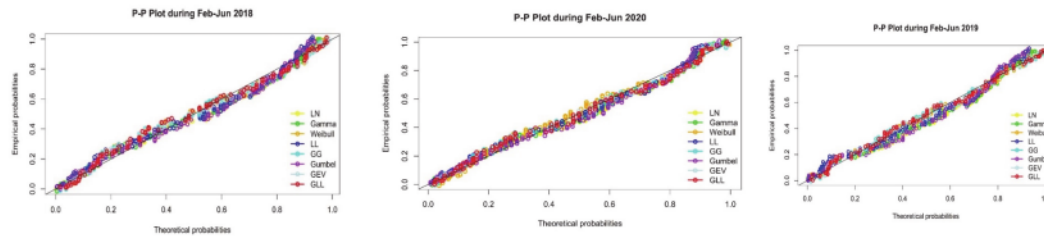


**Fig. 6**. The probability–probability (P–P) plot of the LN, gamma, Weibull, LL, Gumbel, GG, GEV, and GLL distributions of $PM_{2.5}$ concentrations in the period without and during the period with COVID-19.

Therefore, from the standpoint of values of K–S, CvM, A–D, and Log-L values, the generalized distributions seem to be better than the classical distributions, including the LN distribution. The LN distribution is very well known in fitting data of environmental pollution. Since distributions of environmental pollution data are generally positive as well as highly skewed, the LN model is an ideal descriptor of such data, with a positively skewed, positive range, and heavy right tail. Having developed physical mechanisms generating environmental quality data, Ott (1990) provided an argument as to why the LN distribution is so ubiquitous in environmental phenomena. Ott's explanations involve the central limit theorem and the diffusion law. Unfortunately, as it can be seen from histograms, distributions of data of $PM_{2.5}$ concentrations in the periods of February–June of 2018 and 2019 (period without COVID-19) are not positive-skewed distributions. For the data of the period with COVID-19 (February–June 2020), the data histogram shows that the shape is just slightly positively skewed, with a short tail. But in fitting the data of $PM_{2.5}$ concentrations during the period with COVID-19 (February–June 2020), the performance of the LN distribution is still worse than those of other distributions. Furthermore, it can be seen in Table 2, on the basis of almost all goodness-of-fit statistics, the LN distribution fits worse than the other classical distributions. In fact, in terms of all of the proposed goodness-of-fit statistics, the GLL distribution is considerably better than the classical distributions, including the LN distribution, in describing data of $PM_{2.5}$ concentrations in the periods without COVID-19 and during the period with the COVID-19 pandemic in Jakarta, Indonesia. The GLL distribution is an extension of the LL distribution, which is roughly similar in shape to the LN distribution. As noted by Singh *et al.* (1988), the family of four-parameter GLL distributions is a flexible distribution containing several submodels, such as LL, LN, Weibull, and GG. Consequently, the GLL distribution as a general model should provide at least as good a fit as that of other special models. In support of this argument, Warsono *et al.* (2000) demonstrates that the GLL model provides a better representation of environmental pollutant data than do the exponential, LN, Weibull, and GG models.

Furthermore, in the periods of February–June 2018 and February–June 2019, Table 2 shows that, in general, the K–S, CvM, and A–D values of the GLL distributions are smaller than those of the GG and GEV distributions. The log-L value of the GEV distribution is somewhat larger than that of the GLL and GG distributions, but the discrepancy of the values of log-L among distributions is quite small. Thus, in the periods of February–June 2018 and February–June 2019, or the periods without COVID-19, the GLL distribution seems to be an ideal descriptor of such data of $PM_{2.5}$ concentrations. Similar results are obtained for the period of February–June 2020, or the period with the COVID-19 pandemic, for the data of $PM_{2.5}$ concentrations. The Log-L and A–D values of the GEV distribution are slightly higher than those of the GLL distribution, but the discrepancy of the log-L and A–D values between both GEV and GLL distributions is small. However, the K–S and CvM values of the GLL distribution and the GEV distribution are the same. Therefore, the GLL and GEV distributions fit the data in modeling data of $PM_{2.5}$ concentrations during the period with COVID-19. Since the number of parameters of the GLL distribution is higher than that of the other classical distributions, its PDF becomes more mathematically complicated. To estimate the parameters of the GLL distribution, one may involve intensive computations. Fortunately, however, the computational aspect of modeling univariate distributions especially for environmental data is not so difficult because of the availability of the package of R. We have addressed the ability of this package to accommodate general distributions, specifically for distributions not yet defined as functions such as the GLL distribution. In the future, modeling with other general distributions will be possible because of the development of classical distribution models in the case of air pollution with simple computing problems.

## 4. Conclusions

Graphically and numerically, the performances of the generalized distributions are considerably better than those of the classical distributions in fitting the data in the periods without and that during COVID-19 in Jakarta, Indonesia. Particularly, on the basis of the graphs of empirical PDFs of compared models superimposed on relative frequency histograms of data,

the GLL, GEV, and GG distributions seem to be better than the LN, gamma, Weibull, LL, and Gumbel distributions in modeling $PM_{2.5}$ concentrations in the first period without COVID-19 (February–June 2018) and during the period with COVID-19 (February–June 2020). However, in the second period without COVID-19 (February–June 2019), the GLL and GG distributions are the most appropriate models in describing data. The CDF graphs also show that the GLL and GEV distributions perform better than the LN, gamma, Weibull, LL, Gumbel, and GG distributions in the periods without and during the period with the COVID-19 pandemic. Furthermore, according to the P–P plots, the GLL distribution is the most suitable model in representing the data. In short, graphically, the GLL distribution seems to be a promising distribution in modeling the data of $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia. In terms of the values of K–S, CvM, and A–D statistics, in the periods without COVID-19 (February–June 2018 and February–June 2019), the GLL distribution seems to be an ideal descriptor of such data of $PM_{2.5}$ concentrations. Similar results were obtained for the period of February–June 2020, or the period during which the COVID-19 pandemic is present, for the data of $PM_{2.5}$ concentrations. From the standpoint of the log-L values, the GLL model, as well as the GG and GEV distributions, fits. Therefore, on the basis of goodness-of-fit statistics, the GLL distribution is a good alternative to the classical distributions, including the LN distribution, in fitting data of $PM_{2.5}$ concentrations in the periods without COVID-19 (February–June 2018 and February–June 2019) and during the period with COVID-19 (February–June 2020) in Jakarta, Indonesia. In short, the GLL distribution may become a good alternative to the LN distribution and other distributions in fitting data of $PM_{2.5}$ concentrations.

## Acknowledgement

## References

Aguiar-Gil, D., Gómez-Peláez, L.M., Álvarez-Jaramillo, T., Correa-Ochoa, M.A. & Saldarriaga-Molina, J.C. (2020). Evaluating the impact of $PM_{2.5}$ atmospheric pollution on population mortality in an urbanized valley in the American tropics. *Atmospheric Environment, 224*, 117343.

Bencala, K. E., & Seinfeld, J. H. (1976). On frequency distributions of air pollutant concentrations. *Atmospheric Environment (1967), 10*(11), 941-950.

Cameletti, M. (2020). The effect of corona virus lockdown on air pollution: Evidence from the City of Brescia in Lombardia region (Italy). *Atmospheric Environment, 239*, 117794.

Delignette-Muller, M. & Dutang, C. (2015). Fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software, 64*(4), 1-34.

Government of the Republic of Indonesia. (2020). Government Regulation of the Republic of Indonesia No 21 of 2020.

Indonesian Department of Meteorology, Climatology and Geophysics. (2020). Air quality historical data. https://aqicn.org/city/indonesia/jakarta/us-consulate/central/

Kalpasanov, Y. & Kurchatova, G. (1976). A study of the statistical distribution of chemical pollutants in air. *Journal of the Air Pollution Control Association, 26*(10), 981-985.

Kao, A.S. & Friedlander, S.K. (1995). Frequency distributions of PMl0 chemical components and their sources. *Environmental Science and Technology, 29*, 19-28.

Karaca, F., Alagha, O. & Erturk, F. (2004). Statistical characterization of atmospheric $PM_{10}$ and $PM_{2.5}$ concentrations at a non-impacted suburban site of Istanbul, Turkey. *Chemosphere, 59*, 1183–1190.

Kerimray, A., Baimatova, N., Ibragimova, O.P., Bukenov, B., Kenessov, B., Plotitsyn, P. & Karaca, F. (2020). Assessing air quality changes in large cities during COVID-19 lockdowns: The impacts of traffic-free urban conditions in Almaty, Kazakhstan. *Science of The Total Environment, 730*, 139179.

Kushner, E.J. (1976). On determining the statistical parameters for pollution concentration from a truncated data set. *Atmospheric Environment, 10*, 975-979.

Lavagnini, I. & Camuffo, D. (1987). Applications of the Generalized gamma distribution to air pollution problems: The case of Venice. *Science of The Total Environment 61*, 91-105.

Li, L., Li , Q., Huang, L., Wang, Q., Zhu, A., Xu, J., Liu, Z., Li, H., Shi, L., Li, R., Azari, M., Wang, Y., Zhang, X., Liu, Z., Zhu, Y., Zhang, K., Xue, S., Ooi, M.C.G., Zhang, D. & Chan, A. (2020). Air quality changes during the COVID-19 lockdown over the Yangtze River Delta Region: An insight into the impact of human activity pattern changes on air pollution variation. *Science of The Total Environment 732*, 132282.

Li, T, Hu, R., Chen, Z., Li, Q., Huang, S., Zhu, Z. & Zhou, L.F. (2018). Fine particulate matter ($PM_{2.5}$): The culprit for chronic lung diseases in China. *Chronic Disease and Translational Medicine, 4*, 176-186.

Lu, H.C. (2002). The statistical characters of $PM_{10}$ concentration in Taiwan area. *Atmospheric Environment, 36*, 491-502.

Lu, H.C. & Fang, G.C. (2002). Estimating the frequency distributions of PM and PM by the 10 2.5 statistics of wind speed at Sha-Lu, Taiwan. *Science of The Total Environment 298*, 119–130.

Mage, D.T. & Ott, W.R. (1984). An evaluation of the methods of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. *Atmospheric Environment, 18*(1), 163-171.

Marani, A., Lavagni, I. & Buttazzoni, C. (1986). Statistical study of air pollutant concentrations via generalized gamma distributions. J. Air Poll. Control Assoc. *36*(11), 1250-1254.

Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*(4), 308–313.

Otmani, A., Benchrif , A., Tahri, M., Bounakhla, M., Chakir, E.M., Bouch, M.E. & Krombi, M. (2020). Impact of Covid-19 lockdown on $PM_{10}$, $SO_2$ and $NO_2$ concentrations in Salé City (Morocco). *Science of The Total Environment 735*, 139541.

Ott, W.R. (1990). A Physical Explanation of the Lognormality of Pollutant Concentrations. *Journal of the Air and Waste Management, 40*(10), 1378-1383.

Owen, W.J. & De Rouen, T.A. (1980). Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. *Biometrics, 36*, 707-719.

Plocoste, T., Calif, R., Euphrasie-Clotilde, L. & Brute, F. (2020). The statistical behavior of PM10 events over guadeloupean archipelago: Stationarity, modelling and extreme events. *Atmospheric Research, 241*, 104956.

Singh, K.P., Lee, C.M.S. & George, E.O. (1988). On generalized log-logistic model for censored survival data. *Biometrical Journal, 7*, 843-850.

Warsono, Usman, M. & Nusyirwan. (2000). On the estimation of the generalized log-logistic distribution with applications to pollutant concentration data. *Forum Statistika dan Komputasi,* Special Edition, 74-77. (in Indonesia).

WHO (World Health Organization). (2020). WHO announces COVID-19 outbreak a pandemic. https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic.

Xi, W., Jie, C.R., Heng, C.B. & Dong, K.H. (2013). Application of Statistical Distribution of $PM_{10}$ Concentration in Air Quality Management in 5 Representative Cities of China. *Biomedical and Environmental Science, 26*(8), 638-646.

Zambrano-Monserrate, M.A., Ruano, M.A. & Sanchez-Alcalde, L. (2020). Indirect effects of COVID-19 on the environment. *Science of The Total Environment, 728*, 138813.

# Modeling generalized statistical distributions of PM2.5 concentrations during the COVID-19 pandemic in Jakarta, Indonesia

method", Atmospheric Environment (1967), 1982

Publication

---

5    Okur, M.C.. "On fitting the generalized @l-distribution to air pollution data", Atmospheric Environment (1967), 1988    1 %

Publication

---

6    digitalcommons.wayne.edu    1 %

Internet Source

---

7    Thomas Plocoste, Rudy Calif, Lovely Euphrasie-Clotilde, Francenor Brute. "The statistical behavior of PM10 events over guadeloupean archipelago: Stationarity, modelling and extreme events", Atmospheric Research, 2020    1 %

Publication

---

| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |

# Modeling generalized statistical distributions of PM2.5 concentrations during the COVID-19 pandemic in Jakarta, Indonesia

FINAL GRADE

/0

GENERAL COMMENTS

**Instructor**

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8