Journal of Applied Mathematics and Computational Mechanics 2026, 25(1), 7-18
www.amcm.pcz.pl p-ISSN 2299-9965
DOI: 10.17512/jamcm.2026.1.01 e-ISSN 2353-0588

COMPARATIVE ANALYSIS OF EXTREME VALUE DISTRIBUTIONS FOR MODELING PARTICULATE MATTER AIR POLLUTION IN TRENČÍN, SLOVAKIA

Ivana Pobočíková[1], Mária Michalková[1], Zuzana Sedliačková[1], Daniela Jurášová[2] Branislav Ftorek[1]

[1] *Department of Applied Mathematics, University of Žilina, Žilina, Slovakia*
[2] *Department of Building Engineering and Urban Planning, University of Žilina, Žilina, Slovakia*
ivana.pobocikova@fstroj.uniza.sk, maria.michalkova@fstroj.uniza.sk
zuzana.sedliackova@fstroj.uniza.sk, daniela.jurasova@uniza.sk, branislav.ftorek@fstroj.uniza.sk

Received: 19 September 2025; Accepted: 28 November 2025

Abstract. Long-term exposure to high levels of PM_{10} and $PM_{2.5}$ represents a serious health risk and causes significant environmental damage. Therefore, it is necessary to monitor and predict the concentrations of air pollutants to implement both short-term and long-term actions to prevent health-risky situations. This study aims to determine the optimal probability distribution for modeling PM_{10} and $PM_{2.5}$ concentrations in the city of Trenčín, Slovakia. The dataset, comprising daily average concentrations of PM_{10} and $PM_{2.5}$ measured from January 1, 2017, to December 31, 2024, was fitted with the lognormal, gamma, log-logistic, Weibull, and exponentiated Weibull distributions. The exponentiated Weibull probability distribution has not previously been applied to such a dataset. To identify the best-fitting distribution, five performance indicators were employed: the root mean squared error, the coefficient of determination, the prediction accuracy, the Kolmogorov-Smirnov and Anderson-Darling tests. The Weibull distribution generally exhibited the worst performance. In contrast, the exponentiated Weibull distribution consistently ranked among the top three best distributions.

MSC 2010: *62P12, 92F05*
Keywords: *PM_{10}, $PM_{2.5}$, urban traffic, probability distribution fitting, exponentiated Weibull*

1. Introduction

Intensive human activities, including industrial processes, agriculture, transportation, and the generation of heat and electricity, are major sources of environmental pollution. Air pollution represents one of the most significant environmental health risks affecting the population of Europe. The main harmful air pollutants include particulate matter (PM), ground-level ozone (O_3), ammonia (NH_3), nitrogen dioxide (NO_2), nitrogen monoxide (NO), carbon monoxide (CO), sulfur dioxide (SO_2), and volatile organic compounds (VOC) [1]. These pollutants are known to cause a range of health problems, such as cardiovascular and respiratory problems.

PM refers to fine solid or liquid particles found in the air. They are distinguished by their diameter: PM_{10} is defined as particles with a nominal diameter of 10 micrometers (μm) or less, while $PM_{2.5}$ is defined as particles with a diameter of 2.5 micrometers (μm) or less. Due to their minute size, particulate matter can be transported by wind over long distances and remain suspended in the atmosphere for extended periods [2]. This prolonged atmospheric presence contributes to adverse effects on both human health and the environment. The European Air Quality Directive 2008/50/EC [3] sets the limit value for PM_{10} at 40 $\mu g/m^3$ for the annual average and 50 $\mu g/m^3$ for the daily average, with a maximum number of exceedances set at 35 days per year. The limit value for $PM_{2.5}$ is set at 20 $\mu g/m^3$ for the annual average.

Particulate matter can originate from a wide range of sources, both natural and anthropogenic. In Slovakia, the primary source of PM_{10} emissions is household heating, particularly the combustion of solid fuels, which accounts for over 60 % of total PM_{10} emissions. Road transport contributes less than 10 %, while industrial and energy-related sources are responsible for approximately 10 %. Emissions from waste management and agriculture contribute to a lesser extent. Similarly, in the case of $PM_{2.5}$, household heating, again dominated by the use of solid fuels, is the leading emission source, contributing up to 80 % of total annual emissions [4].

In this study, we aim to model particulate matter (PM_{10} and $PM_{2.5}$) concentrations in the city of Trenčín (48°53' N, 18°02' E). This city, located in northwestern Slovakia, is the administrative center of a self-governing region and an important center of trade, economy, culture, and sports. The city's territory is located on the border of two basins, surrounded on the west and east by mountains. Trenčín lies in the northern temperate zone and has a continental climate with four distinct seasons. The prevailing winds are western and northwestern.

Given the high population density and significant levels of industrial, transport, and construction activity, the territory of Trenčín has been included in the air quality management system. The city is among the areas with a medium pollution load, primarily from transport, with a lessening contribution from industrial production. Currently, the pollution from energy sources is decreasing, but that from special production and transport is increasing [5].

Measurements of pollutant concentrations in the air are carried out and reported by the Slovak Hydrometeorological Institute (SHMU). To facilitate comprehensive air quality monitoring, a national air quality monitoring network (NMSKO) has been established. In Trenčín, an automatic air quality monitoring station (AMS) was established at Hasičská Street (longitude 18°02'29'' E, latitude 48°53'47'' N, altitude 214 m). The station is located near a large intersection with high traffic intensity. Therefore, it is classified as an urban traffic station (based on the predominant emission source and the building density). The station measures the gaseous pollutants (SO_2, benzene, NO, NO_2, NO_x, CO) and particulate matter (PM_{10} and $PM_{2.5}$). The measured meteorological parameters include wind speed and direction, temperature, and air humidity. Sampling of PM_{10} and $PM_{2.5}$ is conducted at a height of 4 m above the ground and 1.2 m above the roof of the container using the TEOM 1405F device. Sampling frequency during the continuous measurement of PM_{10} and $PM_{2.5}$ is set at 15 minutes. The automatic monitoring device provides hourly average concentrations of PM_{10}, $PM_{2.5}$, SO_2, benzene, NO, NO_2, NO_x, CO and ozone [4].

Particulate matter concentrations can be considered random variables because they vary over time and space, influenced by various unpredictable factors, such as natural and anthropogenic emission sources, local meteorological conditions, and geography [6]. Modeling PM concentrations using appropriate probability distribution is essential for understanding their behavior in the atmosphere. A properly selected probability distribution enables more accurate statistical analysis and forecasting, which is crucial for effective ambient air quality monitoring. Moreover, such modeling supports the development of effective and efficient policies and measures in air quality management [7]. Numerous probability distributions have been successfully applied to fit PM concentration data. Among the most commonly used is the lognormal distribution [6, 8, 9]. Other widely applied distributions include the Weibull [10, 11], gamma [12, 13], and log-logistic distributions [14, 15]. Additionally, more complex models such as the Burr and Burr-Weibull mixture [16], Pearson type IV and V [17, 18], and the q-exponential [19] and q-Weibull distributions [20] have also shown utility in specific contexts. However, there is no universal probability distribution suitable for all datasets. The choice of an appropriate distribution depends on several factors, such as the nature of the dataset, the processes generating the data, the type of pollutant, the geographical location, the measurement period, and the averaging time used [10].

Given that PM concentration data consist of strictly positive values and typically exhibit a positively skewed distribution with a longer tail on the right, we compared the fit of the following probability distributions: two-parameter lognormal, three-parameter Weibull, three-parameter gamma, and two-parameter log-logistic. In addition to these, the exponentiated Weibull distribution was also fitted. This distribution, although utilized in fields like wind speed [21], reliability analysis, finance, and medicine [22], had not yet been applied to PM pollutants.

2. Methods

The cumulative distribution functions $F(x)$ and probability density functions $f(x)$ of the fitted probability distributions are summarized in Table 1. For all of them, the random variable $x > 0$ or $x > c \geq 0$ when a positive location parameter is present.

Let $\hat{F}(x)$ be the estimated cumulative distribution function. Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be the observations in an ascending order, i.e., $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. The empirical distribution function $F_n(x)$ is the average of the values of the indicator function $I(x_{(i)} \leq x)$, $i = 1, \dots, n$, where $I(x_{(i)} \leq x) = 1$ if $x_{(i)} \leq x$ and 0 otherwise.

Table 1. The cumulative distribution function (CDF) and the probability density function (PDF) of the fitted probability distributions

Probability distribution		Cumulative distribution function $F(x)$ Probability density function $f(x)$	Parameters
Weibull	CDF	$F(x) = 1 - \exp\left(-\left(\frac{x-c}{b}\right)^a\right)$	$a > 0$ – shape parameter $b > 0$ – scale parameter $c > 0$ – location parameter
	PDF	$f(x) = \frac{a}{b^a}(x-c)^{a-1}\exp\left(-\left(\frac{x-c}{b}\right)^a\right)$	
Gamma	CDF	$F(x) = \frac{\gamma\left(\alpha, \frac{x-c}{\beta}\right)}{\Gamma(\alpha)}$	$a > 0$ – shape parameter $b > 0$ – scale parameter $c > 0$ – location parameter
	PDF	$f(x) = \frac{1}{\Gamma(\alpha)\,\beta^\alpha}(x-c)^{\alpha-1}\exp\left(-\frac{x-c}{\beta}\right)$	$\gamma(p,x) = \int_0^x e^{-t}\,t^{p-1}dt$ – lower incomplete Gamma function ($p > 0$)
Lognormal	CDF	$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\mu \in R$ – location parameter $\sigma \geq 0$ – scale parameter
	PDF	$f(x) = \frac{1}{x\,\sigma\sqrt{2\,\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)$	$\Phi(x)$ – CDF of the standard normal distribution
Log-logistic	CDF	$F(x) = \frac{e^z}{1+e^z}$	$\mu > 0$ – location parameter $\sigma > 0$ – scale parameter
	PDF	$f(x) = \frac{e^z}{\sigma x(1+e^z)^2}$	$z = \frac{\ln(x) - \mu}{\sigma}$
Exponentiated Weibull	CDF	$F(x) = \left[1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)\right]^\gamma$	$a > 0$ – shape parameter $b > 0$ – scale parameter $\gamma > 0$ – shape parameter
	PDF	$f(x) =$ $= \frac{\gamma a}{b}\left(\frac{x}{b}\right)^{a-1}\left[1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)\right]^{\gamma-1}\cdot$ $\cdot \exp\left(-\left(\frac{x}{b}\right)^a\right)$	

For evaluating the goodness-of-fit of the distributions, two tests are applied: the Kolmogorov-Smirnov (*KS*) and the Anderson-Darling (*AD*) test. In the *KS* test, the null hypothesis, which states that the data follow the distribution being tested, is rejected at the significance level α if the test statistic

$$D = \max_{1 \leq i \leq n}\left[\left|\hat{F}\left(x_{(i)}\right) - \frac{i-1}{n}\right|, \left|\frac{i}{n} - \hat{F}\left(x_{(i)}\right)\right|\right] \tag{1}$$

obtains a value higher than the critical value $D(\alpha)$ of the *KS* test. The *AD* test gives more weight to the tails of the distribution than the *KS* test does. The null hypothesis of the *AD* test, stating that data follow the specified distribution, is rejected at the significance level α if the test statistic

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln\left(\hat{F}\left(x_{(i)}\right)\right) + \ln\left(1 - \hat{F}\left(x_{(n+1-i)}\right)\right) \right] \quad (2)$$

is greater than the critical value of the *AD* test. When comparing two probability disributions, the one with the smaller test statistic indicates a better fit to the data.

The error is measured by the root mean squared error

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} \left[F_n(x_i) - \hat{F}(x_i) \right]^2 \right)^{\frac{1}{2}}. \quad (3)$$

As other performance indicators, the coefficient of determination is used

$$R^2 = \frac{\sum_{i=1}^{n} \left[\hat{F}(x_i) - \bar{F} \right]^2}{\sum_{i=1}^{n} \left[\hat{F}(x_i) - \bar{F} \right]^2 + \sum_{i=1}^{n} \left[F_n(x_i) - \hat{F}(x_i) \right]^2}, \quad (4)$$

as well as the prediction accuracy [12]

$$PA = \sum_{i=1}^{n} \frac{\left(\hat{F}(x_i) - \bar{F} \right)\left(F_n(x_i) - \bar{F}_n \right)}{(n-1)\sigma_p \sigma_o}. \quad (5)$$

Here $\bar{F}$ stands for the average of $\hat{F}(x_i)$, and $\bar{F}_n$ stands for the average of $F_n(x_i)$; σ_p is the standard deviation of the predicted values, σ_o is the standard deviation of the observed values. The values for these two indicators range from 0 to 1. The probability distribution that obtains values closer to one for both R^2 and PA, better fits to the data.

3. Results

We modeled two datasets: one containing the hourly average concentrations of PM_{10} and the other containing the hourly average concentrations of $PM_{2.5}$, both in μg/m^3. Initially, both these datasets comprised 70,128 data points, covering 2,922 days from January 1, 2017, to December 31, 2024. These hourly data were then converted into daily averages. To ensure data quality, the daily average was assigned only if at least 75 % of the hourly data (18 records) were available for that day. Following this rule, the dataset of daily average PM_{10} concentrations ultimately contained 2,864 data points (98.02 %), while the $PM_{2.5}$ dataset consisted of 2,862 data points (97.95 %). Daily averages were further split into datasets by years.

The descriptive statistics for the daily concentrations of both PM_{10} and $PM_{2.5}$ are presented in Table 2. For PM_{10}, the highest annual mean concentration was observed in 2017, while the lowest was recorded in 2023. None of the annual average concentrations exceeded the limit value of 40 μg/m^3. However, the situation with the daily averages differs. Comparing the maximal value of the daily averages to the limit of 50 μg/m^3, it is evident that this threshold was exceeded at least once each year. The numbers of days exceeding the limit concentration are as follows: 41 days (2017), 39 days (2018), 24 days (2019), 17 days (2020), 16 days (2021), 8 days (2022), 6 days (2023), and 16 days (2024). The highest daily concentration recorded during the entire period was 127.503 μg/m^3, occurring on April 1, 2024. This peak was attributed to Saharan dust, with similarly high concentrations observed across the country.

The coefficient of variation exceeds 49 %, indicating a very high variability in PM_{10} concentrations at this location. Positive skewness, ranging from 1.066 to 2.530 across all modeled years, implies a strongly right-skewed distribution. Positive kurtosis values suggest a leptokurtic distribution compared to the normal distribution, particularly in 2024, when the distribution should be considered highly leptokurtic.

Table 2. Descriptive statistics of the data per type and year

Daily average concentrations of PM_{10}								
Year	Mean [μg/m^3]	Standard deviation [μg/m^3]	Min [μg/m^3]	Max [μg/m^3]	Skewness	Kurtosis	Median [μg/m^3]	Coefficient of variation [%]
2017	29.731	17.554	2.420	106.636	1.550	3.463	26.100	59.042
2018	29.386	15.257	4.516	85.793	1.246	1.642	25.295	51.919
2019	24.946	15.178	4.271	98.836	1.781	4.252	20.489	60.844
2020	23.576	12.134	1.845	65.389	1.066	1.048	21.151	51.466
2021	27.041	13.288	6.297	101.110	1.555	4.587	24.300	49.141
2022	22.559	11.382	4.851	68.793	1.142	1.402	19.971	50.456
2023	18.955	10.678	3.414	76.596	1.813	5.400	16.300	56.335
2024	21.859	13.223	2.198	127.503	2.530	13.043	18.754	60.495
Daily average concentrations of $PM_{2.5}$								
Year	Mean [μg/m^3]	Standard deviation [μg/m^3]	Min [μg/m^3]	Max [μg/m^3]	Skewness	Kurtosis	Median [μg/m^3]	Coefficient of variation [%]
2017	13.220	9.845	1.432	40.553	0.873	–0.020	11.030	74.476
2018	19.689	11.717	1.876	70.236	1.467	2.513	16.487	59.509
2019	17.693	12.281	2.670	83.997	2.102	5.790	13.853	69.413
2020	14.535	9.309	0.403	48.336	0.893	0.592	12.991	64.047
2021	14.766	9.968	0.410	57.301	0.991	1.051	12.794	67.507
2022	13.698	9.365	1.320	57.370	1.484	2.788	11.171	68.364
2023	11.937	8.472	0.834	64.937	2.210	7.776	9.899	70.970
2024	13.276	8.659	0.567	62.156	1.591	4.135	11.726	65.223

For $PM_{2.5}$, none of the annual average concentrations exceeded the limit value of 20 μg/m^3. The highest annual mean concentration 19.689 μg/m^3 was observed in 2018. The daily averages for $PM_{2.5}$ exhibit higher variability than PM_{10}, with the coefficient of variation exceeding 59 %. The data for every year suggest a right-skewed and leptokurtic probability distribution, with one exception. In 2017, the underlying distribution should be considered platykurtic. This year also shows the highest variability in the data (74.5 %), indicating that the data are less concentrated around the mean.

The maximum likelihood method was chosen to estimate the parameters of the probability distributions due to its favorable asymptotic properties, including efficiency and consistency. For more information on this estimation method, see, for example [9]. Table 3 presents a summary of the estimated parameters.

Table 3. Estimated parameters of the distributions for the PM_{10} and $PM_{2.5}$ datasets

Concentrations of PM_{10}														
Year	Weibull			Gamma			Lognormal		Log-logistic		Exponentiated Weibull		
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{a}$	$\hat{b}$	$\hat{\gamma}$
2017	1.66	30.87	2.24	2.85	9.97	1.30	3.23	0.59	3.25	0.33	0.95	12.88	4.49
2018	1.74	28.23	4.31	3.12	8.40	3.20	3.25	0.51	3.26	0.29	0.88	9.39	7.89
2019	1.47	23.06	4.15	2.26	9.45	3.57	3.06	0.56	3.05	0.32	0.66	3.91	14.47
2020	1.92	24.89	1.54	3.89	6.06	0	3.03	0.54	3.04	0.30	1.18	13.84	3.35
2021	1.65	23.42	6.14	2.86	7.69	5.04	3.19	0.47	3.19	0.27	0.90	8.44	9.06
2022	1.64	19.98	4.73	2.60	7.05	4.19	3.00	0.50	3.00	0.29	0.86	6.54	9.17
2023	1.56	17.45	3.32	2.55	6.33	2.78	2.80	0.53	2.81	0.30	0.77	4.42	10.66
2024	1.62	22.15	2.08	2.98	6.96	1.10	2.93	0.57	2.94	0.31	0.88	7.75	6.01
Concentrations of $PM_{2.5}$													
Year	Weibull			Gamma			Lognormal		Log-logistic		Exponentiated Weibull		
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{a}$	$\hat{b}$	$\hat{\gamma}$
2017	1.11	12.22	1.43	1.13	10.48	1.43	2.26	0.86	2.30	0.52	1.04	10.07	1.64
2018	1.63	20.15	1.70	2.87	6.58	0.79	2.82	0.59	2.82	0.33	0.89	7.54	5.22
2019	1.35	16.52	2.64	1.90	8.01	2.51	2.68	0.61	2.67	0.34	0.51	1.13	27.46
2020	1.57	16.01	0.14	2.11	6.89	0	2.42	0.81	2.50	0.44	1.67	16.89	0.93
2021	1.46	16.08	0.17	1.87	7.90	0	2.40	0.88	2.49	0.47	1.66	18.03	0.84
2022	1.39	13.65	1.28	1.87	6.70	1.17	2.40	0.69	2.41	0.39	0.75	4.17	5.42
2023	1.44	12.34	0.80	2.15	5.24	0.67	2.27	0.66	2.28	0.37	0.69	2.75	7.44
2024	1.56	14.27	0.48	2.50	5.30	0	2.37	0.70	2.41	0.39	1.05	8.06	2.57

Based on the results of the goodness-of-fit tests and the model selection criteria, the probability distributions were ranked from the best (1st) to the worst (5th) for respective years (Tables 4 and 5).

Table 4. The goodness-of-fit tests and model selection criteria for the PM_{10} concentration datasets

Year	Probability distribution	R^2	*RMSE*	*PA*	*KS*-test	*AD*-test	Rank
2017	Weibull	0.9902	0.0270	0.9975	0.0541	1.8032	5th
	Gamma	0.9977	0.0136	0.9992	0.0352	0.4693	3rd
	Lognormal	0.9951	0.0198	0.9985	0.0472	0.7981	4th
	Log-logistic	0.9986	0.0110	0.9993	0.0267	0.4053	2nd
	Exponentiated Weibull	0.9988	0.0099	0.0995	0.0260	0.2399	1st
2018	Weibull	0.9862	0.0325	0.9949	0.0590	2.3459	5th
	Gamma	0.9949	0.0204	0.9977	0.0439	0.9246	4th
	Lognormal	0.9983	0.0119	0.9993	0.0330	0.4082	1st
	Log-logistic	0.9979	0.0135	0.9990	0.0357	0.4381	2nd
	Exponentiated Weibull	0.9978	0.0134	0.9989	0.0314	0.4420	3rd
2019	Weibull	0.9779	0.0404	0.9923	0.0759	3.2486	5th
	Gamma	0.9878	0.0309	0.9951	0.0697	1.7859	4th
	Lognormal	0.9948	0.0206	0.9976	0.0525	0.7949	3rd
	Log-logistic	0.9967	0.0167	0.9984	0.0492	0.5920	1st
	Exponentiated Weibull	0.9953	0.0195	0.9978	0.0510	0.7250	2nd
2020	Weibull	0.9869	0.0316	0.9953	0.0621	2.2586	5th
	Gamma	0.9961	0.0178	0.9983	0.0402	0.7308	3rd
	Lognormal	0.9942	0.0213	0.9985	0.0455	0.9914	4th
	Log-logistic	0.9993	0.0078	0.9997	0.0229	0.3248	1st
	Exponentiated Weibull	0.9972	0.0151	0.9988	0.0341	0.5394	2nd
2021	Weibull	0.9955	0.0186	0.9989	0.0435	0.8679	5th
	Gamma	0.9987	0.0101	0.9995	0.0287	0.2802	2nd
	Lognormal	0.9984	0.0114	0.9994	0.0265	0.3414	3rd
	Log-logistic	0.9981	0.0129	0.9992	0.0364	0.4935	4th
	Exponentiated Weibull	0.9991	0.0087	0.9996	0.0236	0.2248	1st
2022	Weibull	0.9940	0.0219	0.9975	0.0517	0.9607	5th
	Gamma	0.9978	0.0136	0.9989	0.0327	0.3764	4th
	Lognormal	0.9984	0.0117	0.9993	0.0304	0.3626	2nd
	Log-logistic	0.9982	0.0124	0.9993	0.0333	0.5529	3rd
	Exponentiated Weibull	0.9985	0.0113	0.9993	0.0264	0.3453	1st
2023	Weibull	0.9891	0.0283	0.9972	0.0597	1.8081	5th
	Gamma	0.9967	0.0162	0.9989	0.0395	0.5905	4th
	Lognormal	0.9986	0.0108	0.9994	0.0310	0.2497	2nd
	Log-logistic	0.9986	0.0111	0.9994	0.0351	0.2529	3rd
	Exponentiated Weibull	0.9988	0.0097	0.9995	0.0289	0.2076	1st
2024	Weibull	0.9779	0.0391	0.9953	0.0781	3.5488	5th
	Gamma	0.9930	0.0230	0.9980	0.0492	1.2550	4th
	Lognormal	0.9938	0.0218	0.9984	0.0513	1.1433	3rd
	Log-logistic	0.9987	0.0104	0.9994	0.0276	0.3412	1st
	Exponentiated Weibull	0.9963	0.0170	0.9989	0.0444	0.7255	2nd

Table 5. The goodness-of-fit tests and model selection criteria for the $PM_{2.5}$ concentration datasets

Year	Probability distribution	R^2	*RMSE*	*PA*	*KS*-test	*AD*-test	Rank
2017	Weibull	0.9893	0.0317	0.9960	0.0624	2.2270	1st
	Gamma	0.9886	0.0325	0.9956	0.0689	2.2052	2nd
	Lognormal	0.9784	0.0456	0.9910	0.0899	4.5478	5th
	Log-logistic	0.9803	0.0430	0.9907	0.0785	4.7942	4th
	Exponentiated Weibull	0.9878	0.0340	0.9953	0.0718	2.9436	3rd
2018	Weibull	0.9797	0.0389	0.9929	0.0702	3.2939	5th
	Gamma	0.9903	0.0278	0.9960	0.0585	1.6436	4th
	Lognormal	0.9956	0.0185	0.9986	0.0449	0.9077	2nd
	Log-logistic	0.9981	0.0126	0.9991	0.0300	0.4292	1st
	Exponentiated Weibull	0.9953	0.0195	0.9979	0.0423	0.8643	3rd
2019	Weibull	0.9685	0.0477	0.9895	0.0967	4.5783	5th
	Gamma	0.9808	0.0384	0.9927	0.0840	2.8527	4th
	Lognormal	0.9923	0.0248	0.9966	0.0576	1.2115	3rd
	Log-logistic	0.9963	0.0175	0.9982	0.0385	0.7731	1st
	Exponentiated Weibull	0.9945	0.0210	0.9975	0.0491	0.8958	2nd
2020	Weibull	0.9991	0.0087	0.9996	0.0264	0.2060	2nd
	Gamma	0.9956	0.0192	0.9985	0.0473	0.7124	3rd
	Lognormal	0.9711	0.0476	0.9907	0.0834	4.8508	5th
	Log-logistic	0.9899	0.0291	0.9949	0.0573	2.8694	4th
	Exponentiated Weibull	0.9993	0.0078	0.9996	0.0255	0.1870	1st
2021	Weibull	0.9966	0.0167	0.9985	0.0409	0.6164	2nd
	Gamma	0.9928	0.0244	0.9974	0.0533	1.2348	3rd
	Lognormal	0.9599	0.0559	0.9866	0.1029	6.5612	5th
	Log-logistic	0.9884	0.0308	0.9943	0.0644	3.7371	4th
	Exponentiated Weibull	0.9971	0.0155	0.9986	0.0390	0.4798	1st
2022	Weibull	0.9928	0.0239	0.9970	0.0605	1.1434	5th
	Gamma	0.9966	0.0167	0.9984	0.0470	0.5227	4th
	Lognormal	0.9979	0.0133	0.9992	0.0348	0.4604	3rd
	Log-logistic	0.9981	0.0128	0.9992	0.0410	0.5920	2nd
	Exponentiated Weibull	0.9984	0.0117	0.9992	0.0307	0.3199	1st
2023	Weibull	0.9855	0.0325	0.9964	0.0610	2.4854	5th
	Gamma	0.9948	0.0202	0.9983	0.0534	0.9092	4th
	Lognormal	0.9974	0.0145	0.9990	0.0365	0.3739	3rd
	Log-logistic	0.9987	0.0104	0.9994	0.0373	0.2902	1st
	Exponentiated Weibull	0.9986	0.0106	0.9994	0.0274	0.2201	2nd
2024	Weibull	0.9952	0.0193	0.9987	0.0470	1.1816	4th
	Gamma	0.9985	0.0109	0.9994	0.0311	0.4239	2nd
	Lognormal	0.9903	0.0278	0.9970	0.0582	1.6735	5th
	Log-logistic	0.9955	0.0195	0.9978	0.0405	1.1221	3rd
	Exponentiated Weibull	0.9986	0.0106	0.9994	0.0267	0.3639	1st

4. Conclusion

Effective air quality management requires robust statistical tools for informed decision-making, risk assessment, and strategy implementation. Probability distributions are fundamental to understanding air pollutant behavior. Selecting the correct distribution is crucial for accurate modeling and reliable conclusions, making it an effective tool for improving air quality.

This paper identified the most suitable probability distribution for average daily PM_{10} and $PM_{2.5}$ concentrations in Trenčín, Slovakia, a city with intensive traffic as a primary pollution source. As noted in [16], when the background atmosphere is strongly influenced by anthropogenic pollution sources, the lognormal distribution has proved optimal for high PM_{10} concentrations. In addition to the lognormal distribution, we tested other probability distributions commonly used to model PM concentrations: the Weibull, gamma, and log-logistic distributions. The aim of the paper was to compare their performance to that of the exponentiated Weibull distribution.

Based on the criteria (R^2, *RMSE*, *PA*) and goodness-of-fit tests (*KS*-test, *AD*-test), the three-parameter Weibull distwribution achieved the worst results among the tested distributions, though its fit to the data can be considered good. This is primarily because pollution concentration data typically exhibit high positive kurtosis, which the Weibull distribution struggles to accommodate. However, when the empirical distribution had kurtosis values closer to zero, the Weibull distribution ranked best or second best in our models ($PM_{2.5}$ dataset in 2017, 2020, and 2021). In the majority of the modeled datasets, the exponentiated Weibull and log-logistic distributions demonstrated the best fit. For the exponentiated Weibull distribution, the following observations may be made:

- It ranked first four times in both the $PM_{2.5}$ and PM_{10} datasets and was second best twice in the $PM_{2.5}$ dataset and three times in the PM_{10} dataset. On the other hand, the log-logistic distribution ranked first only three times in both the $PM_{2.5}$ and PM_{10} datasets.
- It never fell below third place overall.
- It provided the best fit for the $PM_{2.5}$ concentrations in 2020 and 2021, where the data exhibited moderate right skewness and moderate positive kurtosis. However, it also ranked first in years when the data were highly right-skewed and strongly leptokurtic (e.g., PM_{10} in 2021 or 2023, $PM_{2.5}$ in 2024).
- It outperformed the lognormal distribution in both PM datasets in all years except 2018.

Overall, we can conclude that the exponentiated Weibull distribution may be considered an adequate (if not superior) substitute for the commonly used lognormal distribution when modeling PM concentrations. Its second shape parameter (the exponent) allows it to effectively accommodate data with varying degrees of right skewness and kurtosis, making it a valuable tool for air quality analysis.

Acknowledgement

We would like to express our gratitude to SHMU, particularly to Mrs. Jana Matejovičová for data processing, her invaluable advice and willingness to help, which facilitated the writing of this paper. This work has been supported by the Slovak Grant Agency KEGA through the project No. 025ŽU-4/2024 Implementation of new didactic tools to increase the quality of mathematics teaching in the engineering degree at technical universities.

References

[1] EEA (2023). Air Pollution in Europe: 2023 Reporting Status Under the National Emission Reduction Commitments Directive. Publications Office, LU. Available at https://data.europa.eu/doi/10.2800/339055.

[2] Wilson, W.E., & Suh, H.H. (1997). Fine particles and coarse particles: Concentration relationships relevant to epidemiologic studies. *Journal of the Air & Waste Management Association*, *47*, 12, 1238-1249.

[3] The European Parliament and the Council of the European Union (2008). Directive 2008/50/EC on ambient air quality and cleaner air for Europe. Available at https://eur-lex.europa.eu/eli/dir/2008/50/oj.

[4] SHMU (2024). 2023 Air Pollution in the Slovak Republic (report). Available online: https://www.shmu.sk/File/oko/rocenky/2023_Air_Pollution_in_the_SR_v1.pdf.

[5] Stratégia adaptability mesta Trenčín na klimatickú zmenu (2022). Available at https://trencin.sk/pre-obcanov/investicie/projekty-eu/strategia-adaptability-na-klimaticku-zmenu/.

[6] Lu, H.C. (2002). The statistical characters of PM_{10} concentration in Taiwan area. *Atmospheric Environment*, *36*(3), 491-502.

[7] Tran, P.T.M., Adam, M.G., Tham, K.W., Schiavon, S., Pantelic, J., Linden, P.F., Sofianopoulou, E., Sekhar, S.C., Cheong, D.K. W., & Balasubramanian, R. (2021). Assessment and mitigation of personal exposure to particulate air pollution in cities: An exploratory study. *Sustainable Cities and Society*, *72*, 103052.

[8] Pietruczuk, A., & Jarosławski, J. (2013). Analysis of particulate matter concentrations in Mazovia region, central Poland, based on 2007-2010 data. *Acta Geophysica*, *61*(2), 445-462.

[9] Anand, A., Garg, V.K., Agrawal, A., & Pathak, A. (2024). Distribution and concentration pathway of particulate pollution during pandemic-induced lockdown in metropolitan cities in India. *International Journal of Environmental Science and Technology*, *21*, 1993-2006.

[10] Georgopoulos, P.G., & Seinfeld, J.H. (1982). Statistical distributions of air pollution concentrations. *Environmental Science & Technology*, *16*, 7.

[11] Pachauri, R.K., Singh, V., & Dixit, S. (2024). Identifying optimal statistical distributions for air pollution data of Agra. *Journal of Computational Analysis & Applications*, *33*(8), 1109.

[12] Noor, N.M., Tan, C.Y., Ramli, N.A., Yahaya, A.S., & Yusof, N.F.F.M. (2011). Assessment of various probability distributions to model PM_{10} concentration for industrialized area in Peninsula Malaysia: A case study in Shah Alam and Nilai. *Australian Journal of Basic and Applied Sciences*, *5*(12), 2796-2811.

[13] Meirelles, M.G., & Vasconcelos, H.C. (2025). Physical-statistical characterization of PM_{10} and $PM_{2.5}$ concentrations and atmospheric transport events in the Azores during 2024. *Earth*, *6*, 54.

[14] Karaca, F., Alagha, O., & Ertürk, F. (2005). Statistical characterization of atmospheric PM_{10} and $PM_{2.5}$ concentrations at a non-impacted suburban site of Istanbul, Turkey. *Chemosphere*, *59*(8), 1183-1190.

[15] Jiang, X., Deng, S., Liu, N., & Shen, B. (2011). The statistical distributions of SO_2, NO_2 and PM_{10} concentrations in Xi'an, China. *2011 International Symposium on Water Resource and Environmental Protection, Xi'an*, 2206-2212.

[16] Plocoste, T., Calif, R., Euphrasie-Clotilde, L., & Brute, F.N. (2020). Investigation of local correlations between particulate matter (PM_{10}) and air temperature in the Caribbean basin using Ensemble Empirical Mode Decomposition. *Atmospheric Pollution Research*, *11*(10), 1692-1704.

[17] Gavriil, I., Grivas, G., Kassomenos, P., Chaloulakou, A., & Spyrellis, N. (2006). An application of theoretical probability distributions, to the study of PM_{10} and $PM_{2.5}$ time series in Athens, Greece. *Global NEST Journal*, *8*(3), 241-251.

[18] Mijić, Z., Tasić, M., Rajšić, S., & Novaković, V. (2009). The statistical characters of PM_{10} in Belgrade area. *Atmospheric Research*, *92*(4), 420-426.

[19] He, H., Schäfer, B., & Beck, C. (2022). Spatial heterogeneity of air pollution statistics in Europe. *Scientific Reports*, *12*, 12215.

[20] Sánchez, E. (2023). Q-Weibull distribution to explain the $PM_{2.5}$ air pollution concentration in Santiago de Chile. *The European Physical Journal B*, *96*, 108.

[21] Akgül, F.G., & Şenoğlu, B. (2019). Comparison of wind speed distributions: a case study for Aegean coast of Turkey. *Energy Sources A: Recovery, Utilization, and Environmental Effects*, *45*(1), 2453-2470.

[22] Al-Hussaini, E.K., Ahsanullah, M. (2015). *Exponentiated Distributions* (vol. 5). Atlantis Press.