Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 5, October, 2025

Extreme Value Model to Forecast PM2.5 Concentration Through a Non-Stationary Process

Pannarat Guayjarernpanishk [1], Tawun Remsungnen [1], Nipaporn Chutiman [2], Monchaya Chiangpradit [2], Butsakorn Kong-ied [2*]

[1] *Faculty of Interdisciplinary Studies, Nong Khai Campus, Khon Kaen University, Nong Khai 43000, Thailand.*

[2] *Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand.*

Abstract

The objectives of this research were to develop a model to forecast and estimate the return levels for daily maximum PM2.5 concentrations in Thailand, applying Extreme Value Theory (EVT) with the Generalized Extreme Value (GEV) distribution under eight models for stationary and non-stationary process. This research utilized reanalysis data from the NASA EARTHDATA satellite, represented as grid points with a spatial resolution of 50 × 62.5 km, enabling the analysis of daily maximum PM2.5 concentrations across 176 grid points from January 1, 2009 to October 31, 2024. The analysis revealed that Model 2 ($\mu(t) = \beta_0 + \beta_1 t$ where σ and ξ are constants) is the most suitable model for five grid points, namely Sa Kaeo Province, Uthai Thani Province, Nakhon Ratchasima Province, Bueng Kan Province and Mae Hong Son Province, whereas Model 1 (μ, σ and ξ are constants) is suitable for the remaining 171 grid points. Estimating the return levels for return periods of 5, 10, 25, and 50 years showed that Northern Thailand had the most extreme daily PM2.5 concentrations, for all return periods especially Mae Hong Son Province. The results of this analysis can serve as valuable information to support decision-making for response planning in high-risk areas, aiding in efficient resource allocation and preventive measures.

Keywords:

Extreme Value Model;

Non-Stationary;

PM2.5 Concentrations.

Article History:

Received:	08	April	2025
Revised:	09	August	2025
Accepted:	16	August	2025
Published:	01	October	2025

1- Introduction

The concentration of greenhouse gases in the Earth's atmosphere has risen rapidly, particularly since the Industrial Revolution. Over the last 50 years, human activities have driven carbon dioxide levels up from 280 ppm (parts per million), a value that had remained stable for millions of years. By 2022, the concentration had reached 421 ppm, representing a 50% increase compared to pre-industrial levels. This significant rise has exacerbated global warming, causing the climate to increasingly deviate from its original state. Thailand is inevitably affected by global warming. Over the past decade, Thailand's overall air quality has steadily deteriorated, with many areas continuing to experience air pollutant levels that exceed standard limits. One of the primary pollutants remains "particles measuring 2.5 microns or smaller." The growing environmental and pollution issues, along with the continued rise in greenhouse gas emissions, present a significant challenge to achieving sustainable economic growth in the country.

Many researchers have developed models to forecast PM2.5 concentrations, including Sudumbrekar et al. [1], which developed an effective model to forecast PM2.5 in India using the ARIMA model. In 2022, Zhao et al. [2] studied the forecasting of Beijing's PM2.5 using a hybrid ARIMA model. Sudha & Suguna [3] presented an ARIMA model for

* **CONTACT**: butsakorn.k@msu.ac.th

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-05-025

PM2.5 forecasting in Chennai, India. Amelia et al. [4] forecasted PM2.5 pollution in Jakarta using exponential smoothing and ARIMA forecasting methods, and Wang et al. [5] established a monitoring and forecasting system for PM2.5 and other real-time environmental data in China's opencast coal mines using ARIMA and Double Exponential Smoothing models. In 2023, Duan et al. [6] built an air quality forecasting model using the ARIMA model. In 2024, Hernández et al. [7] examined the influence of human activities on urban air quality, and used the ARIMA model to examine the impact of COVID-19 isolation measures on PM10 and PM2.5 levels in an upland Latin American city, Bogotá, Colombia. Pyae & Kallawicha [8] studied the distribution of air pollutants, including PM2.5, PM10, and O3, using multiple linear regression modeling, which also included a calculation of the Air Quality Index (AQI) and the development of an ARIMA model to predict the AQI of PM2.5 and PM10 in Myanmar. Gao et al. [9] analyzed and predicted the concentrations of air pollutants (PM2.5, PM10, SO2, and CO) and the atmospheric environmental quality in Hunan Province, China, using the ARIMA model. Mahawan et al. [10] studied a Situation and temporal behaviors of air pollution in Chiang Mai, Thailand by ARIMA model. Sharma et al. [11] presented a model to systematically forecast PM2.5 concentrations in India. Abuouelezz, et al. [12] explored PM2.5 and PM10 ML forecasting models in a comparative study at six ground stations in Abu Dhabi, United Arab Emirates. Nourmohammad & Rashidi [13] analyzed ground data analysis for PM2.5 prediction using predictive modeling techniques.

Nonetheless, the estimation and forecasting of PM2.5 using the ARIMA model remains limited to short-term forecasts. Therefore, spatially accurate forecasting especially, for long-term predictions is both important and challenging, enabling relevant agencies to plan for long-term responses.

Many researchers have applied extreme value theory to develop models for forecasting long-term extreme air pollution for example, Pornsopin et al. [14] studied risk analysis of PM2.5 at Khon Kaen city, Thailand. Intarapak and Supapakorn [15] investigated the statistical distribution of PM2.5 concentration in Chiang Mai, Thailand. Bodhisuwan & Aryuyuen [16] utilized the poisson transmuted Janardan distribution for modelling count data. Aguirre et al. [17] developed a novel tree ensemble model to approximate the generalized extreme value distribution parameters of the PM2.5 maxima in the Mexico City metropolitan area. Peter et al. [18] studied trends of extreme events and long-term health impacts of particulate matter in a southern Indian industrial area. Yang et al. [19] investigated extreme event discovery with self-attention for PM2.5 anomaly prediction. Klinjan et al. [20] analyzed extreme value with new generalized extreme value distributions for risk analysis on PM2.5 and PM10 in Pathum Thani, Thailand. Vazquez et al. [21] studied bivariate analysis of pollutants monthly maxima in Mexico City using extreme value distributions and copula. Guayjarernpanishk et al. [22] developed statistical model of air pollution forecasting in a regional context for sustainable management. Ghosh [23] assessed air quality extremes via extreme value analysis of metropolitan cities across India and the world. Maricq & Bishop [24] conducted an extreme value theory analysis of high emitter trends across four US cities from 1995 to 2021.

From the research mentioned above, the data used for analysis was collected from air pollution quality measurements in each country and satellite data. Extreme value theory under a stationary process was applied in the data analysis method used in the aforementioned research. In general, when analyzing extreme data, weather data may follow a non-stationary process, involving other variables such as time or season. Therefore, before analyzing the data to determine the model parameters, it is crucial to assess whether the data follows a stationary or non-stationary process. Each process requires a different analysis procedure and methods to select the suitable models. Failing to consider the data characteristics may result in an inaccurate model parameter estimation, which can negatively impact subsequent applications and lead to serious consequences, particularly in data analysis requiring high model accuracy. Considering the significance of the aforementioned issues, there is an urgent need for forecasting air pollution-related information. Consequently, this study aimed to develop a long-term forecasting model for extreme PM2.5 concentrations using a non-stationary process for regions in Thailand. Satellite data was utilized to predict potential extreme PM2.5 concentrations, providing relevant agencies with valuable insights to enhance their operational efficiency.

2- Scope of Research

2-1- Study Area

Thailand lies between latitudes 5°37′N and 20°27′N and longitudes 97°22′E and 105°37′E. The country is categorized into six regions (Figure 1): the northern region (9 provinces in orange), the central region (22 provinces in yellow), the northeastern region (20 provinces in pink), the eastern region (7 provinces in purple), the western region (5 provinces in green), and the southern region (14 provinces in blue).

- Northernmost point: Latitude 20°27'30"N, Mae Sai District, Chiang Rai Province.
- Southernmost point: Latitude 5°37′N, Betong District, Yala Province.
- Easternmost point: Longitude 105°37'30"E, Phibun Mangsahan District, Ubon Ratchathani Province.
- Westernmost point: Longitude 97°22'E, Mae Lan Noi District, Mae Hong Son Province.

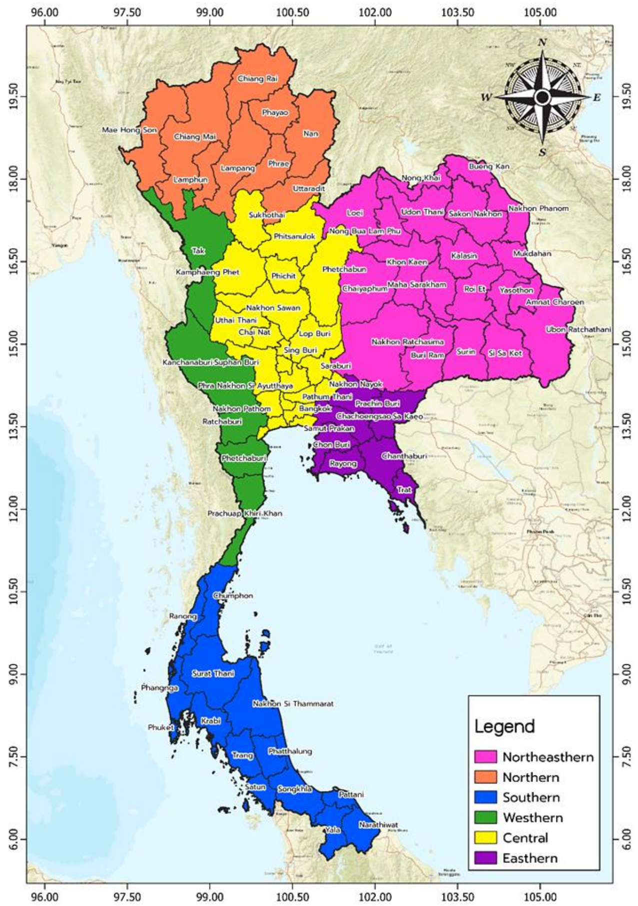


Figure 1. **Study area covering the 6 regions of Thailand**

2-2- Data Preparation

Data obtained from the NASA EARTHDATA satellite (https://giovanni.gsfc.nasa.gov/giovanni/) consisted of grid point data with a spatial resolution of 50 x 62.5 km and hourly dust concentration data. In this research, the daily maximum PM2.5 concentrations from January 1, 2009 to October 31, 2024, totaling 176 grid points, were used, which consisted of 35 grid points in the northern region, 26 in the central region, 53 in the north-eastern region, 12 in the eastern region, 17 in the western region, and 33 in the southern region.

3- Methodology

3-1- Generalized Extreme Value (GEV) Distribution

Jenkinson (1955) [25] developed a method for analyzing extreme event known as the Generalized Extreme Value (GEV) distribution. The concept behind this method of extreme value analysis involves dividing the data into block time, with each block time having an equal duration, such as weekly, monthly, quarterly, or yearly. The maximum or minimum value from each block time is then analyzed as follows:

The concept of a case in which the extreme value used in GEV analysis is the maximum value:

Let X_i ($i = 1, 2, \ldots, n$) denote independent and identically distributed random variables with distribution function, $F(x)$, and define $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$. In GEV analysis, three parameters are being considered: μ (location parameter), σ (scale parameter), and ξ (shape parameter).

Brockett & Galambos (1980) [26] presented the Cumulative Distribution Function (CDF) of the GEV when $-\infty < x < \infty$ as follows:

$$G(x;\mu,\sigma,\xi) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{1}$$

As defined on $\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$, when $-\infty < \mu, \xi < \infty$ and $\sigma > 0$.

If $\xi = 0$ or$\xi \to 0$, we get

$$G(x;\mu,\sigma,\xi) = exp\left\{-\,exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \tag{2}$$

Considering Equation 1, in the case where $\xi < 0$, the Generalized Extreme Value distribution is referred to as the "Weibull Distribution". If $\xi > 0$, it is referred to as the "Fréchet Distribution". Additionally, considering Equation 2, in the case where$\xi \to 0$, it is called the "Gumbel Distribution". The GEV has the following Probability Distribution Function (pdf):

$$g(x;\mu,\sigma,\xi) = \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}; \xi \neq 0 \\ \frac{1}{\sigma} exp\left(\frac{x-\mu}{\sigma}\right)^{-1} exp\left\{-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}; \xi = 0 \end{cases} \tag{3}$$

The concept of the case where the extreme values used in GEV analysis are minimum values is considered in contrast to the case where the extreme values are maximum values.

3-2- Extreme Values under a Non-stationary Process

The analysis of data under a non-stationary process is a comprehensive approach that encompasses all conditions of distribution models, including a stationary process. Therefore, extreme values used for analysis should be examined under a non-stationary process, such as data related to meteorology, hydrology, finance, insurance, or economics. An example of data characteristics under a non-stationary process is illustrated in Figure 2.

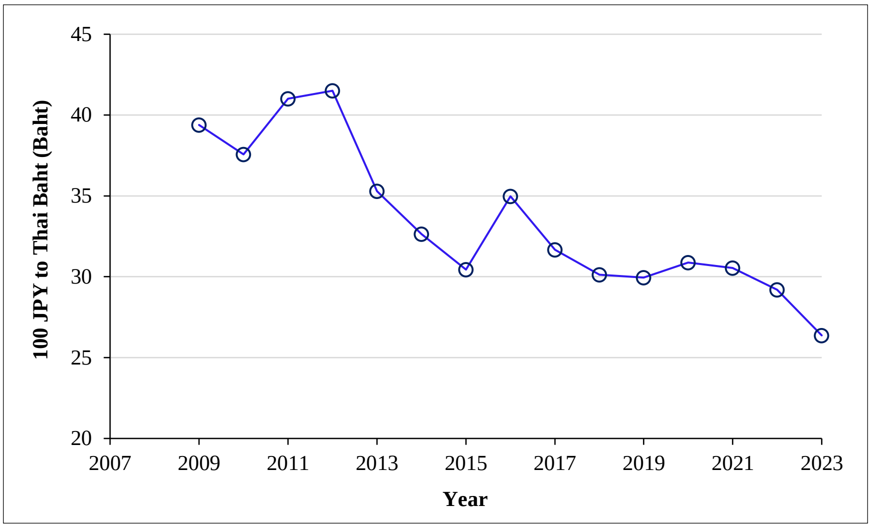


Figure 2. Maximum daily exchange rate JPY to Thai Baht between 2009 and 2023

The GEV distribution model under a non-stationary process, which describes the distribution of X_t at time t (where $t = 1, 2, ..., m$), can be represented as follows:

$$X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t))$$

Parameters within a non-stationary process include, for example:

$X_t \sim \text{GEV}(\beta_0 + \beta_1 t, \sigma, \xi)$, $X_t \sim \text{GEV}(exp(\beta_0 + \beta_1 t), \sigma, \xi)$, $X_t \sim \text{GEV}(\beta_0 + \beta_1 t, exp(\beta_0 + \beta_1 t), \xi)$,$X_t \sim \text{GEV}(\mu, \sigma, \beta_0 + \beta_1\, exp(-\beta_2 t))$ and $X_t \sim \text{GEV}(\beta_0 + \beta_1 t + \beta_2 t^2, \sigma, \beta_0 + \beta_1 t)$ where β_0, β_1, β_2 μ, σ and ξ are parameters.

The possible parameter models are as follows:

Model 1 : $\mu, \sigma$and ξ are constants.

Model 2 : $\mu(t) = \beta_0 + \beta_1 t$, where σ and ξ are constants.

Model 6 : $\mu(t) = \beta_0 + \beta_1 t$ and $\sigma(t) = exp(\beta_0 + \beta_1 t)$, where ξ is a constant.

Model 3 : $\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$, where σ and ξ are constants.

Model 4 : $\mu(t) = \beta_0 + \beta_1\, exp(-\beta_2 t)$, where σ and ξ are constants.

Model 5 : $\sigma(t) = exp(\beta_0 + \beta_1 t)$, where μ and ξ are constants.

Model 7 : $\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ and $\sigma(t) = exp(\beta_0 + \beta_1 t)$, where ξ is a constant.

Model 8 : $\mu(t) = \beta_0 + \beta_1\, exp(-\beta_2 t)$ and $\sigma(t) = exp(\beta_0 + \beta_1 t)$, where ξ is a constant.

If the parameter analysis results align with Model 1, the data used for the analysis can be deemed stationary. Conversely, if the results correspond to other models, the data can be categorized as non-stationary.

3-3- Parameter Estimation

The process of estimating GEV distribution parameters using the maximum likelihood estimation (MLE) method involves the following steps:

- Build the likelihood function of the GEV probability distribution function, which will give:

  $L(\beta)=\prod_{t=1}^{m} g(x_t; \mu(t), \sigma(t), \xi(t)),$

  where β is a vector of parameter β_i, and $g(x_t; \mu(t), \sigma(t), \xi(t))$represents the probability distribution function of the GEV, with $\mu(t)$,$\sigma(t)$ and $\xi(t)$denoting the parameters at x_t.

- Construct the log-likelihood function of the GEV probability distribution derived in step 1 and Equation 3 for t = 1, 2, .., m as follows

  2.1. In the case where;

$$l(\beta) = -\sum_{t=1}^{m}\left\{log\, \sigma(t) + \left(1+\frac{1}{\xi(t)}\right) log\left[1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right)\right] + \left[1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right)\right]^{-\frac{1}{\xi(t)}}\right\} \tag{4}$$

  defined on $1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right) > 0$

  2.2. In the case where $\xi = 0$ or$\xi \to 0$

$$l(\beta) = -\sum_{t=1}^{m}\left\{log\, \sigma(t) + \left(\frac{x_t-\mu(t)}{\sigma(t)}\right) + exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \tag{5}$$

  An example of the GEV distribution under a non-stationary process:

  When $\mu(t) = \beta_0 + \beta_1\, exp(-\beta_2 t)$, with σ and ξ are constants, we get:

$$l(\mu(t), \sigma, \xi) = -\sum_{t=1}^{m}\left\{log\, \sigma + \left(1+\frac{1}{\xi}\right) log\left[1+\xi\left(\frac{x_t-(\beta_0+\beta_1\, e^{-\beta_2 t})}{\sigma}\right)\right] + \left[1+\xi\left(\frac{x_t-(\beta_0+\beta_1\, exp(-\beta_2 t))}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

- Calculate a partial derivative of the functions obtained in step 2 to estimate the parameters $\mu(t)$, $\sigma(t)$, and $\xi(t)$ $\left(\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}(t)\right)$ as follows:

  $\frac{\partial l}{\partial \mu(t)}(\mu(t), \sigma(t), \xi(t)) = 0,$

  $\frac{\partial l}{\partial \sigma(t)}(\mu(t), \sigma(t), \xi(t)) = 0$

  and $\frac{\partial l}{\partial \xi(t)}(\mu(t), \sigma(t), \xi(t)) = 0$

3-4- Model Selection

Selecting the most suitable model for the given data under a non-stationary process using the maximum likelihood estimation method, as discussed in the previous section, means that if more than one related model is obtained, they are referred to as "Nested Models." The statistic used for testing in this case is the Deviance Statistic (*D*). The concept behind the deviance statistic test is based on the following hypotheses:

H_0 :The initial model is appropriate.

H_1 :The model being compared is appropriate.

LetM_0andM_1be the initial model and the model being compared, respectively, under the condition $M_0 \subset M_1$. The statistic *D*, defined as follows, is used:

$$D(0,1) = 2\{l_1(M_1) - l_0(M_0)\}, \tag{6}$$

where $l_0(M_0)$ and $l_1(M_1)$ represent the maximum log-likelihood values of M_0 and M_1, in that order. Based on Equation 6, the statistic *D* converges in distribution to a chi-square distribution with n (χ_n^2) degrees of freedom, where n is the difference in the number of parameters between M_1 and M_0.$M_0$is rejected at the level of significance (α(if $D > c_\alpha$, where c_α is the quantile at $(1-\alpha)$ of χ_n^2. If the null hypothesis is rejected, it indicates that M_1 explains the variability in the data better than M_0. The test using statistic *D* is conducted pairwise, comparing two models at a time. After a suitable model is found, it is tested against other models until the most suitable one is determined.

According to the parameter models in Section 3.2, where M_i represents the models at *i*, the following values are obtained:$M_1 \subset M_3 \subset M_7$,$M_1 \subset M_2 \subset M_6$, $M_1 \subset M_4 \subset M_8$and $M_1 \subset M_5$. Figure 3. shows the steps for the sequential pairwise comparison, while Figure 4. presents the flowchart detailing the pairwise comparison process for the first case)$M_1 \subset M_3 \subset M_7$.(

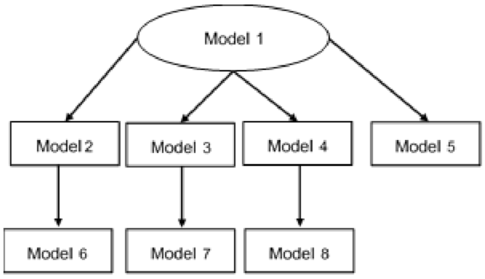


Figure 3. The steps for the sequential pairwise comparison

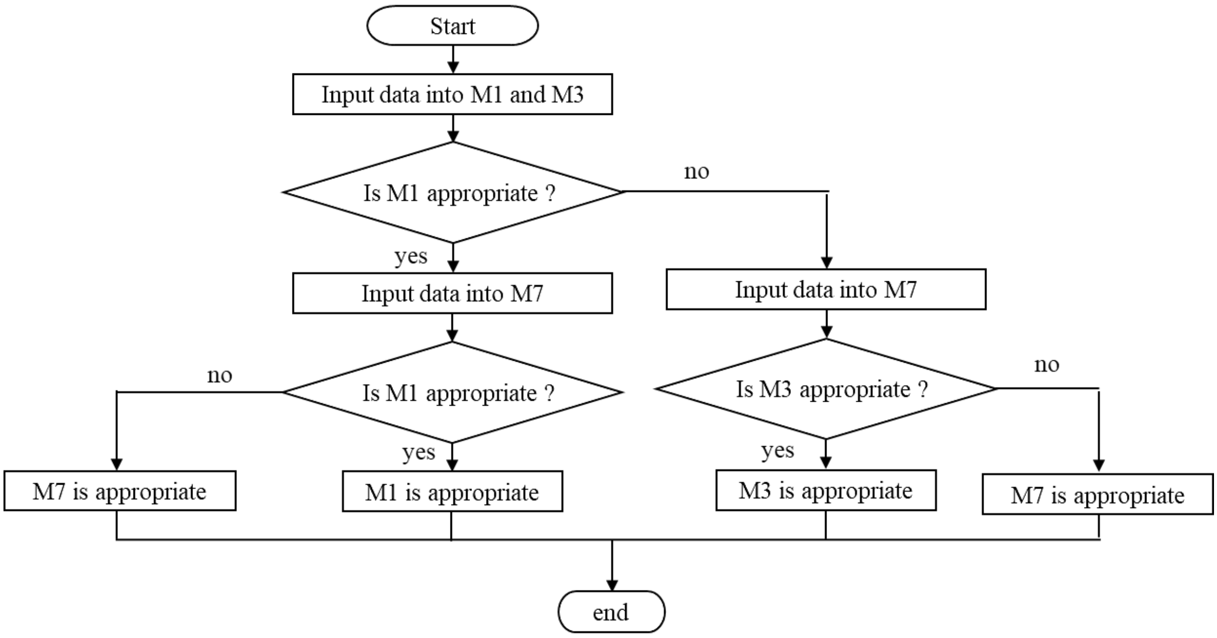


Figure 4. Example of the pairwise comparison process for the case $M_1 \subset M_3 \subset M_7$

After determining the best-fit model, it is used to estimate the return level at time T years ($\hat{R}_T$) using the following formula:

$$\hat{R}_T = \begin{cases} \hat{\mu}(t) - \frac{\hat{\sigma}(t)}{\hat{\xi}(t)}\left\{1 - \left[-\log\left(1 - \frac{1}{T}\right)\right]^{-\hat{\xi}(t)}\right\}; & \xi \neq 0 \\ \hat{\mu}(t) - \hat{\sigma}(t)\log\left\{-\log\left(1 - \frac{1}{T}\right)\right\}; & \xi = 0 \end{cases} \tag{7}$$

4- Results and Discussion

The data used in this study consisted of the daily maximum PM2.5 concentrations from 176 grid points, covering the period from January 2009 to October 2024. The statistical values of the top three daily maximum dust concentrations in each region of Thailand are shown in Table 1.

From Table 1, it is evident that points with the maximum PM2.5 concentrations over the past 15 years, which exceed 50 micrograms per cubic meter and pose health risks, are found in the northern, central, north-eastern, and southern regions. In the western and eastern regions, the maximum PM2.5 concentrations were below 50 micrograms per cubic meter. The western region recorded the lowest maximum PM2.5 concentrations, ranging between 30.195 and 32.640 micrograms per cubic meter. The highest maximum PM2.5 concentrations was recorded in the north-eastern region in Loei Province in June 2010. The area with the second-highest PM2.5 concentrations was in the central region in Nakhon Sawan Province in June 2010. Both grid points exhibited maximum PM2.5 concentrations exceeding 100 micrograms per cubic meter.

In applying extreme value theory for modeling, the Generalized Extreme Value (GEV) distribution is used, setting a monthly block time and analyzing daily maximum PM2.5 concentrations for each month. The analysis employed a method for calculating extreme values under a non-stationary process across eight models ($M_1, M_2, \ldots, M_8$), using data from 176 grid points. The results indicated that Model 2 (M_2) was suitable for 5 grid points, while Model 1 (M_1) was suitable for the remaining 171 grid points. Table 2 presents the estimated parameters for the suitable model at the grid points with the three highest daily maximum PM2.5 concentration values in each region, and Table 3 provides the estimated parameter values for Model 2 at all 5 grid points.

Table 1. **Statistical values of the top two daily maximum PM2.5 concentrations in each region**

Regions	Lat [°N]	Long [°E]	Provinces	Average daily maximum PM2.5 $[\frac{\mu}{m^3}]$	Daily maximum PM2.5 $[\frac{\mu}{m^3}]$	Month/Year of daily maximum PM2.5 $[\frac{\mu}{m^3}]$
Northern	17.50	98.750	Lamphun	9.909	95.315	May 20١٨
	١٩.٥٠	97.500	Mae Hong Son	10.648	91.881	June 2012
	18.00	101.250	Uttaradit	9.755	83.557	June 2021
Central	15.50	100.000	Nakhon Sawan	8.840	102.820	June 2012
	16.00	100.000	Nakhon Sawan	9.009	88.214	June 2012
	15.00	100.000	Chai Nat	8.262	69.442	August 2018
Western	16.00	98.750	Tak	8.375	48.007	May 2023
	16.50	98.750	Tak	9.026	47.483	May 2024
	17.00	98.750	Tak	9.748	45.387	May 2018
Eastern	14.00	101.250	Prachinburi	7.275	32.640	June 2022
	14.00	101.875	Prachinburi	7.193	31.097	May 2023
	13.00	101.250	Rayong	6.623	30.195	July 2009
North-eastern	17.00	101.875	Loei	10.123	١٢٦,٢٥	June 2010
	16.00	104.375	Yasothon	8.414	70.082	May 2019
	15.50	104.375	Yasothon	7.943	62.981	May 2019
Southern	7.00	99.375	Satun	5.256	79.483	July 2011
	7.00	100.000	Satun	5.021	71.857	July 2011
	6.50	100.000	Satun	5.193	71.421	July 2011

Table 2. **The suitable model for daily maximum PM2.5 concentrations at the points with the highest recorded PM2.5 concentrations in each region**

Regions	Lat [°N]	Long [°E]	Provinces	Estimated parameter values		
				μ	σ	ξ
Northern	17.50	98.750	Lamphun	5.43794	3.86993	0.43372
	١٩.٥٠	97.500	Mae Hong Son	5.79684	4.51935	0.39997
	18.00	101.250	Uttaradit	5.44339	3.88786	0.39731
Central	15.50	100.000	Nakhon Sawan	4.95374	3.19005	0.42020
	16.00	100.000	Nakhon Sawan	5.25034	3.39209	0.37786
	15.00	100.000	Chai Nat	4.83485	3.11694	0.39271
Western	16.00	98.750	Tak	5.08091	3.46251	0.31815
	16.50	98.7500	Tak	5.35282	3.72886	0.34448
	17.00	98.750	Tak	5.68061	4.01496	0.36827
Eastern	14.00	101.250	Prachinburi	4.27072	2.99374	0.35536
	14.00	101.875	Prachinburi	4.17133	2.97524	0.36545
	13.00	101.250	Rayong	3.94226	2.76852	0.32392
Northeastern	17.00	101.875	Loei	5.34227	3.69467	0.44548
	16.00	104.375	Yasothon	4.62896	3.33144	0.41728
	15.50	104.375	Yasothon	4.47031	3.25083	0.38407
Southern	7.00	99.375	Satun	1.86574	1.59438	0.65169
	7.00	100.000	Satun	2.16816	1.84739	0.61591
	6.50	100.000	Satun	2.06116	1.82844	0.66314

Table 3. The suitable model for Model 2

Regions	Lat [ºN]	Long [ºE]	Provinces	Estimated parameter values			
				β_0	β_1	σ	ξ
Eastern	14.00	103.125	Sa Kaeo	4.07296	0.00003	3.03538	0.38493
Central	15.50	99.375	Uthai Thani	4.89261	-0.00001	3.14944	0.33314
North-eastern	15.50	102.50	Nakhon Ratchasima	4.76459	0.00001	3.37142	0.38305
North-eastern	18.50	103.125	Bueng Kan	5.20203	-0.00003	3.39126	0.28689
Northern	19.00	98.125	Mae Hong Son	4.91450	0.00427	3.99081	0.41135

Note: Model 2 is$\mu(t) = \beta_0 + \beta_1 t$, where σ and ξ are constants.

Table 2 presents the estimated parameters of the best-fitting model at the grid points with the three highest daily maximum PM2.5 concentrations. At all three locations, Model 1 was identified as the most suitable. Model 1 consists of three parameters $\mu, \sigma$and ξ, all of which are constants. In contrast, Table 3 shows the estimated parameters of Model 2, which was determined to be the best-fitting model for 5 out of the 176 grid points. In Model 2 $\mu(t) = \beta_0 + \beta_1 t$, where σ and ξ are constants.

After the suitable model for each grid point was determined, the resulting model was used for return level estimation at 5, 10, 25, and 50 years. The return levels of the daily maximum PM2.5 concentrations in areas with the highest recorded PM2.5 concentrations in each region, based on Model 1 in Table 2, are shown in Table 4. The return levels for Model 2, as presented in Table 3, are displayed in Table 5.

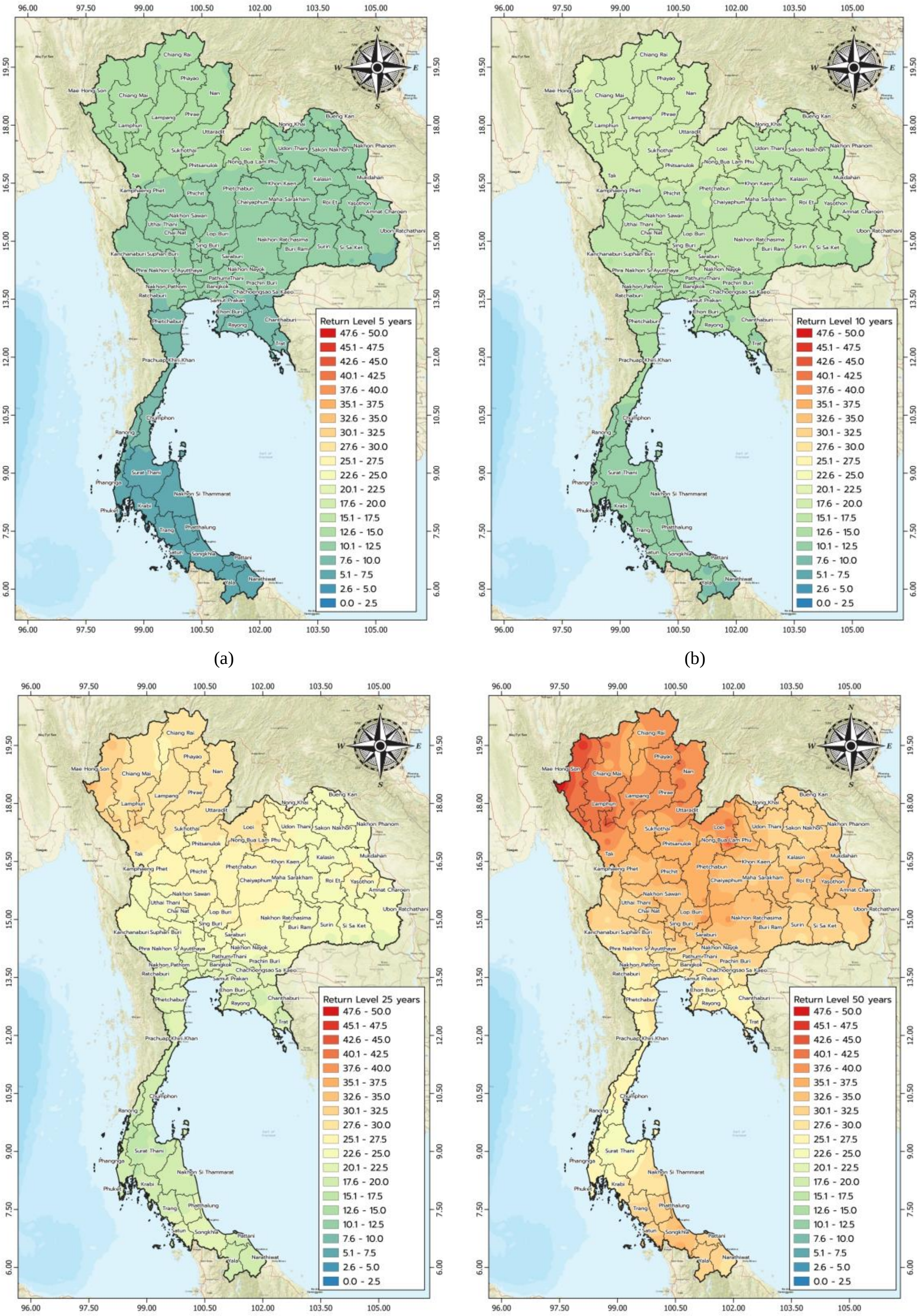
Table 4. Return levels of daily maximum PM2.5 concentrations in areas with the three highest recorded PM2.5 concentrations in each region, according to Model 1 in Table 2

Regions	Lat [ºN]	Long [ºE]	Provinces	Return levels ($\frac{\mu}{m^3}$(			
				5 year	10 year	25 year	50 year
Northern	17.5	98.750	Lamphun	13.61646	20.19510	32.24072	44.98524
	١٩.٥٠	97.500	Mae Hong Son	15.08471	22.29142	35.10912	48.30400
	18.00	101.250	Uttaradit	13.41599	19.58458	30.53086	41.77495
Central	15.50	100.000	Nakhon Sawan	11.62023	16.90587	26.47192	36.48254
	16.00	100.000	Nakhon Sawan	17.28334	26.33577	35.48866	12.09593
	15.00	100.000	Chai Nat	11.20236	16.10484	24.76997	33.63760
Western	16.00	98.750	Tak	16.46634	24.30720	31.85889	11.73674
	16.50	98.7500	Tak	12.67552	18.02893	27.10687	36.03959
	17.00	98.750	Tak	13.71979	19.74927	30.18480	40.65415
Eastern	14.00	101.250	Prachinburi	10.20215	14.58948	22.09903	29.55415
	14.00	101.875	Prachinburi	10.11494	14.55925	22.23262	29.91321
	13.00	101.250	Rayong	9.28897	13.11206	19.48154	25.64502
Northeastern	17.00	101.875	Loei	13.22715	19.64942	31.52845	44.21714
	16.00	104.375	Yasothon	11.57410	17.06356	26.97353	37.31949
	15.50	104.375	Yasothon	11.06437	16.09454	24.91960	33.88774
Southern	7.00	99.375	Satun	6.72407	11.16327	20.67701	32.33961
	7.00	100.000	Satun	6.52492	10.85609	20.14710	31.54667
	6.50	100.000	Satun	6.75903	11.56632	22.29955	35.96659

Table 5. Return levels of daily maximum PM2.5 concentrations, according to Model 2 in Table 3

Regions	Lat [ºN]	Long [ºE]	Provinces	Return levels ($\frac{\mu}{m^3}$)			
				5 year	10 year	25 year	50 year
Eastern	14.00	103.125	Sa Kaeo	10.23964	14.94403	23.20360	31.60291
Central	15.50	99.375	Uthai Thani	11.01954	15.44515	22.87706	30.12285
Northeastern	15.50	102.50	Nakhon Ratchasima	11.59861	16.80522	25.93177	35.19855
Northeastern	18.50	103.125	Bueng Kan	11.55263	15.91921	22.96670	29.58340
Northern	19.00	98.125	Mae Hong Son	14.00467	20.50716	32.18614	44.32038

Tables 4 and 5 show the return levels of PM2.5 concentrations over a time period of T years with a probability of 1/T. For example, a 5-year return level means that within 5 years, PM2.5 concentrations at the predicted level will occur at least once with a probability of 1/5, or 0.2. The predicted values for all 176 grid points at the 5 year, 10 year, 25 year, and 50 year return levels were used to create contour graphs using GIS Kriging interpolation, with the results presented in Figures 5.



(c) (d)

Figure 5. Estimated values for the return levels in (a) 5, (b) 10, (c) 25 and (d) 50 years of daily maximum PM2.5 concentrations

Figure 5 shows that the northern region of Thailand has higher PM2.5 concentrations than the other regions, especially in Mae Hong Son Province and Lamphun Province. Provinces located to the west of the northeastern and central regions have similar PM2.5 concentrations, while the southern region of Thailand has lower PM2.5 concentrations than the other regions.

5- Conclusion

In developing the extreme value theory model using the Generalized Extreme Value distribution under a non-stationary process with 8 models, it is found that Model 2 is the most suitable model for 5 grid points, while the remaining 171 grid points are best fit by Model 1. The return level of daily maximum PM2.5 concentrations increases with longer return periods, which is consistent with the findings of Guayjarernpanishk et al. [22]. The results of the return level estimation for the daily maximum PM2.5 concentrations indicate that the northern region of Thailand has higher PM2.5 concentrations than the other regions, with the values exceeding the WHO recommended air quality guideline of 15 micrograms per cubic meter [27] for all return periods, except for the 5 year return period, primarily due to factors such as wildfires, which are among the most significant contributors. Wildfires can be broadly divided into two types. The first type are wildfires that occur naturally. In the northern region, there are deciduous forests, such as dry dipterocarp forests and mixed deciduous forests, which cover most of the area. When these forests begin shedding their leaves, they create excellent fuel, which can naturally ignite fires. However, natural forest fires are rare, while man-made fires occur more frequently. Another issue is the smog that drifts in from the burning of farmlands in neighboring countries. In the southern region of Thailand, PM2.5 concentrations are lower than in other regions because the wind direction, especially during the southwest and northeast monsoon seasons which help, blow pollution out of the area. Moreover, these winds do not carry as much dust from neighboring countries as those in the northern or north-eastern region. The southern region has a less dense population and fewer industries that emit dust and pollution compared to other regions, such as the northern and north-eastern regions, where burning is used for agriculture, or the central region, where there are many industrial estates.

6- Declarations

6-1- Author Contributions

Conceptualization, P.G. and N.C.; methodology, P.G., M.C., and B.K.; software, P.G.; validation, M.C., N.C., T.R., and B.K.; formal analysis, P.G.; investigation, M.C. and N.C.; resources, B.K.; data curation, B.K.; writing—original draft preparation, N.C. and P.G; writing—review and editing, M.C., T.R., and B.K.; visualization, M.C. and B.K.; supervision, P.G. and N.C.; project administration, P.G.; funding acquisition, P.G. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding

This research was supported by the Fundamental Fund of Khon Kaen University and the National Science, Research and Innovation Fund (NSRF), and Mahasarakham University.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Sudumbrekar, A., Kale, R., Kaurwar, T., Mule, V., & Devkar, A. (2021). Feasibility Study of ARIMA Model for PM2.5 Prediction using Real-world Data Gathered from Pune Region. New Frontiers in Communication and Intelligent Systems, 105–111, SCRS Publication, Delhi, India. doi:10.52458/978-81-95502-00-4-13.

[2] Zhao, L., Li, Z., & Qu, L. (2022). Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition. Heliyon, 8(12), 12239. doi:10.1016/j.heliyon.2022.e12239.

[3] Sudha, G., & Suguna, S. (2022). Health Hazard: PM2. 5 Forecast-A Visual Analytic Framework Using ARIMA. International journal of health sciences, 6(S2), 630-642. doi:10.53730/ijhs.v6nS2.5066.

[4] Amelia, R., Guskarnali, Mahardika, R. G., Niani, C. R., & Lewaherilla, N. (2022). Predicting particulate matter PM2.5 using the exponential smoothing and Seasonal ARIMA with R studio. IOP Conference Series: Earth and Environmental Science, 1108(1), 012079. doi:10.1088/1755-1315/1108/1/012079.

[5] Wang, M., Zhang, Q., Tai, C., Li, J., Yang, Z., Shen, K., & Guo, C. (2022). Design of PM2.5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA Mode. PLOS ONE, 17(5), e0267440. doi:10.1371/journal.pone.0267440.

[6] Duan, J., Gong, Y., Luo, J., & Zhao, Z. (2023). Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. Scientific Reports, 13(1), 12127. doi:10.1038/s41598-023-36620-4.

[7] Hernández-Medina, D. S., Zafra-Mejía, C. A., & Rondón-Quintana, H. A. (2024). ARIMA Analysis of PM Concentrations during the COVID-19 Isolation in a High-Altitude Latin American Megacity. Atmosphere, 15(6), 683. doi:10.3390/atmos15060683.

[8] Pyae, T. S., & Kallawicha, K. (2024). First temporal distribution model of ambient air pollutants (PM2.5, PM10, and O3) in Yangon City, Myanmar during 2019–2021. Environmental Pollution, 347, 123718. doi:10.1016/j.envpol.2024.123718.

[9] Gao, W., Xiao, T., Zou, L., Li, H., & Gu, S. (2024). Analysis and Prediction of Atmospheric Environmental Quality Based on the Autoregressive Integrated Moving Average Model (ARIMA Model) in Hunan Province, China. Sustainability (Switzerland), 16(19), 8471. doi:10.3390/su16198471.

[10] Mahawan, N., Charoentrakulpeeti, W., & Knobnob, N. (2024). A Situation and Temporal Behaviors of Air Pollution by ARIMA Model: A Case Study of Chiang Mai. Asian Creative Architecture, Art and Design, 37(1), 1–16. doi:10.55003/acaad.2024.268045.

[11] Sharma, V., Ghosh, S., Mishra, V. N., & Kumar, P. (2025). Spatio-temporal Variations and Forecast of PM2.5 concentration around selected Satellite Cities of Delhi, India using ARIMA model. Physics and Chemistry of the Earth, 138, 103849. doi:10.1016/j.pce.2024.103849.

[12] Abuouelezz, W., Ali, N., Aung, Z., Altunaiji, A., Shah, S. B., & Gliddon, D. (2025). Exploring PM2.5 and PM10 ML forecasting models: a comparative study in the UAE. Scientific Reports, 15(1), 9797. doi:10.1038/s41598-025-94013-1.

[13] Nourmohammad, E., & Rashidi, Y. (2025). Ground data analysis for PM2.5 Prediction using predictive modeling techniques. Journal of Air Pollution and Health, 10(1), 61–82. doi:10.18502/japh.v10i1.18095.

[14] Pornsopin, J., Busababodhin, P., Phoophiwfa, T., Chiangpradit, M., & Guayjarernpanishk, P. (2021). Risk analysis of PM2.5 and PM10: A case study at Khon Kaen City. The Journal of Applied Science, 20(2), 157–172. doi:10.14416/j.appsci.2021.02.012.

[15] Intarapak, S., & Supapakorn, T. (2021). Investigation on the Statistical Distribution of PM2.5 Concentration in Chiang Mai, Thailand. WSEAS Transactions on Environment and Development, 17, 1219–1227. doi:10.37394/232015.2021.17.111.

[16] Bodhisuwan, W., & Aryuyuen, S. (2022). The Poisson-Transmuted Janardan Distribution for Modelling Count Data. Trends in Sciences, 19(5), 1450–1461. doi:10.48048/tis.2022.2898.

[17] Aguirre-Salado, A. I., Venancio-Guzmán, S., Aguirre-Salado, C. A., & Santiago-Santos, A. (2022). A Novel Tree Ensemble Model to Approximate the Generalized Extreme Value Distribution Parameters of the PM2.5 Maxima in the Mexico City Metropolitan Area. Mathematics, 10(12), 2056. doi:10.3390/math10122056.

[18] Peter, A. E., Raj, M., Gangadharan, P., Athira, P., & Nagendra, S. M. S. (2023). Trends, Extreme Events and Long-term Health Impacts of Particulate Matter in a Southern Indian Industrial Area. Water, Air, and Soil Pollution, 234(5), 303. doi:10.1007/s11270-023-06302-y.

[19] Yang, H. C., Yang, M. C., Wong, G. W., & Chen, M. C. (2023). Extreme Event Discovery With Self-Attention for PM2.5 Anomaly Prediction. IEEE Intelligent Systems, 38(2), 36–45. doi:10.1109/MIS.2023.3236561.

[20] Klinjan, K., Sottiwan, T., & Aryuyuen, S. (2024). Extreme Value Analysis with New Generalized Extreme Value Distributions: A Case Study for Risk Analysis on Pm2.5 and Pm10 in Pathum Thani, Thailand. Communications in Mathematical Biology and Neuroscience, 100. doi:10.28919/cmbn/8833.

[21] Vazquez-Morales, J. A., Rodrigues, E. R., & Reyes-Cervantes, H. J. (2024). Bivariate Analysis of Pollutants Monthly Maxima in Mexico City Using Extreme Value Distributions and Copula. Journal of Environmental Protection, 15(07), 796–826. doi:10.4236/jep.2024.157046.

[22] Guayjarernpanishk, P., Chutiman, N., Piwpuan, N., Kong-Ied, B., & Chiangpradit, M. (2024). Air Pollution Forecasting in a Regional Context for Sustainable Management. Emerging Science Journal, 8(5), 2091-2100. doi:10.28991/ESJ-2024-08-05-024.

[23] Ghosh, D. (2025). Assessing air quality extremes: a comparative extreme value analysis of metropolitan cities across India and the world. Environmental Monitoring and Assessment, 197(3), 276. doi:10.1007/s10661-025-13754-8.

[24] Maricq, M. M., & Bishop, G. A. (2025). Extreme value theory analysis of high emitter trends across four US cities from 1995 to 2021. Science of the Total Environment, 958. doi:10.1016/j.scitotenv.2024.177873.

[25] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158–171. doi:10.1002/qj.49708134804.

[26] Brockett, P. L., & Galambos, J. (1980). The Asymptotic Theory of Extreme Order Statistics. Journal of the American Statistical Association, 75(370), 473. doi:10.2307/2287487.

[27] W.H.O. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization (WHO), Geneva, Switzerland.