MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Investigating the Degree of Persistence, Trend and the Best Time Series Forecasting Models for Particulate Matter (PM_{10}) Pollutant Across Malaysia

Lawan Adamu Isma'il[a], Norhashidah Awang[a*], Ibrahim Lawal Kane[b]

[a]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia; [b]Department of Mathematics and Statistics, Umaru Musa Yar'adua University, 2218 Katsina State, Nigeria

Abstract Particulate matter is the most common atmospheric pollutant with some negative consequences on human health, environment, and the ambient air quality. In this study, the concentration of particulate matter in sixty-five air quality monitoring stations across Malaysia during January to December 2018 is analyzed. We investigated the degree of persistence and trend of the particulate matter series and developed a forecasting model using both the autoregressive integrated moving average (ARIMA) and the autoregressive fractionally integrated moving average (ARFIMA) time series methods for each monitoring station separately. Mean absolute deviation (MAD), mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to determine the best fitted model for forecasting each monitoring station. Ljung-Box test of uncorrelated residuals confirmed the adequacy of each of the model. The results confirmed the evidence of transitory form of persistence in the level of particulate matter pollutant at sixty-four monitoring stations while trend increases in seventeen monitoring stations. Forecast error analysis indicates that ARFIMA models performed better than ARIMA models by producing smaller RMSE values in forty-two of the sixty-five monitoring stations. However, the overall result indicates that none of the model could be regarded as universal in forecasting particulate matter concentration, and their performance is independent of the category or location of a given monitoring station.

Keywords: Particulate Matter, Long memory**,** ARIMA, Mann-Kendall Trend, Forecasting.

Introduction

The sources of air pollution are complex, and typically comes from various sources. In Malaysia, before the year 2018, the calculation of Air Pollution Index (API) was based on five major pollutants namely, ground level ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and particulate matter of size 10 microns or less in diameter (PM_{10}) [1]. On the 16th of August 2018, the particulate matter of size 2.5 microns or less in diameter ($PM_{2.5}$) was included in the API calculation as the sixth pollutant. The API is classified according to its range as good, moderate, unhealthy, very unhealthy, and hazardous. The Department of Environment (DOE) which is a unit of the Malaysian Ministry of Environment is saddled with the responsibility of monitoring air quality status across all Malaysian states. To adequately handle these very important responsibilities, the DOE had in the middle of April 2017, increased the number of its air quality monitoring stations to sixty-five continuous monitoring stations. These stations are strategically located in rural, suburban, urban and industrials areas to monitor changes in the ambient air quality. The Malaysian ambient air quality standard for the year 2018 stood at $45\mu g/m^3$ which is less than the world health organization's standard of $50\mu g/m^3$ [1].

Particulate matter of size 10 micron or less in diameter (PM_{10}) is identified as the most visible and comparatively more rampant air pollutant in Peninsular Malaysia that affects human health and

***For correspondence:**
shidah@usm.my

Received: 3 April 2023
Accepted: 21 Sept. 2023

environment [2, 3, 4, 5, 6, 7]. PM_{10} pollutants are microscopic in nature of solid and liquid forms commonly found in the air. Their sources can be natural or through human activities such as in industrial, power generation and other developmental activities, and through emissions from motor vehicles, land clearing open burning and forest fires. These particles can penetrate the upper respiratory tract; while some of these particles can be eliminated from the body by sputum spitting, toxic particles possessed quicker and greater chances of harming individuals. The particles can enter human bronchi and obstruct gas exchange in the lungs, resulting in disorders such as asthma, bronchitis, and cardiovascular diseases [8, 9, 10, 11, 12]. Long-term exposure to these pollutants increases the risk of developing heart, lung, and other respiratory health concerns, which can lead to a significant decrease in life expectancy [13, 14, 15, 16]. Furthermore, particulate matter pollutant is a large contributor to hazy weather [17, 18] which has a significant impact on day-to-day life as well as social production [19, 20]. The presence of PM_{10} on vegetation's surface produces radiative heating, a decline in photosynthesis rate, and changes in decomposition cycles, all of which affects the animal groups [21].

Various studies have been conducted to study the trend and concentration of air pollutants, including the particulate matter across the globe. For example, Li *et al*. [17] studied the concentration of PM_{10} and $PM_{2.5}$ pollutants in relation to the meteorological condition from seven air monitoring stations in Shijiazhuang city of China over a three-month duration. The average daily concentration of $PM_{2.5}$ was $94.45\mu g/m^3$ while that of PM_{10} was calculated at $219.15\mu g/m^3$, which exceeded the WHO tolerance limit of the pollutants. However, they realized significant positive correlation coefficient between the atmospheric pressure and the pollutants, while significant negative correlation was observed between the atmospheric temperature and the particulate matter pollutants.

In India, Naveen and Anu [22] analyzed and forecasted the varying trend of outdoor air quality using the dataset recorded at different monitoring air quality stations in the district of Thiruvananthapuram, Kerala, India. The analysis was conducted using the Seasonal Autoregressive Integrated Moving Average (SARIMA) and ARIMA models, and ARIMA model was observed to have performed better in giving more accurate forecast values. The study noted that the most disturbing and rampant pollutant in the study area was the respirable suspended particulate matter. Mishra and Goyal [23], had earlier reported that air quality of Delhi, India, has deteriorated and falls within the ranges of "bad" and "very bad" categories.

Kliengchuay *et al*. [24] identified the best fitted model for prediction of particulate matter concentration especially the PM_{10} in Thailand's city of Chiang Rai among other models. Hourly air pollution and weather data recorded at two different stations for eight-year period (2011 until 2018) were used to develop four different stepwise multiple linear regression (MLR) models with steps as annual, summer, rainy and winter. The result revealed that daily maximum PM_{10} concentration was observed in the summer season for the two stations while the minimum daily concentration was detected in the rainy season. It also revealed that seasonal variation of PM_{10} is significantly different between the two stations while CO emission was moderately related with the PM_{10} emission during the summer season. The PM_{10} summer model was graded by the authors as the best MLR model to predict PM_{10} concentration during haze episode over the other three stepwise MLR models. In addition, Aladağ [25] forecasted the particulate concentration (PM_{10}) using traditional and hybridized ARIMA model in Turkey's Erzurum city. Twelve-year monthly observations were used with the first ten-year data used for testing the model while the remaining two-year data used to validate and identify the best performing model over the other. It was established that the wavelet-transform hybridized WT-ARIMA outperformed the conventional ARIMA model through RMSE, R^2, IA, MAE and MAPE.

To the best of our knowledge, little or not much has been done to study the persistence behavior or long memory in PM_{10} concentrations across Malaysia. Therefore, the purpose of this paper is to study the statistical issues of long memory and trend in PM_{10} concentrations and to identify the best forecasting time series model for the sixty-five air monitoring stations spread across Peninsular Malaysia, with the view of providing insight on the level of pollutant to public and relevant government agencies.

The long-range dependence (LRD) also known as long memory was first tested in hydrological data by British hydrologist H. E. Hurst [26], and later in econometrics, earth and environmental sciences, network traffic and linguistics. It is based on the Hurst parameter which means the presence of strong connection effect between values at different lags. The LRD statistically refers to the slower or non-exponential decay of the autocorrelation function (ACF) so that the area under the curve of the function is infinite.

Geweke and Porter-Hudak (GPH) as one of the most popular method for estimating the degree of persistence in the time series data, is employed in this paper to check the presence or otherwise of the long memory using the parameter estimator *d* in the ARFIMA (*p,d,q*) model structure. The GPH estimator is based on the regression equation using the periodogram function as an estimate of the spectral density. In addition, the Mann-Kendall test for trend estimation is used to prove if the contaminant concentrations are increasing or decreasing over time. It is a nonparametric test with the null hypothesis of no trend in the series, the test was found very useful by numerous researchers in estimating the statistical trend in hydrological, climatological and air pollutants studies [27, 28, 29]. The remainder of this research paper is categorized as follows: Section 2 describes the study area including geo-locations, nature of data used for the study, presents the methodological flowchart, and gives definitions of some statistical terms. Section 3 gives some descriptive statistics, presents boxplot graphical representation, tables of stationarity test and results of main analyses. Section 4 gives some concluding remarks.

Materials and Methods

Study Area and Data Used

Malaysia is a country located in southeastern Asia. There are thirteen states and three federal territories separated by South China Sea into two regions, Peninsular Malaysia, and East Malaysia. Peninsular Malaysia shares a land and maritime border with Thailand, and maritime borders with Singapore, Vietnam, and Indonesia. East Malaysia shares land and maritime borders with Brunei and Indonesia, and maritime border with Philippines and Vietnam. The country which is having its national capital at Kuala Lumpur city has the total population estimated at over 32 million people, this placed the country as 43rd most populous country in the world. Malaysia is located between latitude 0^0 53' 82.652" north and 7^0 1' 19.014" north of the equator. It is on longitude 101^0 4' 8.444" east and 119^0 6' 722" east of prime meridian. The study utilized a daily data collected over a one-year period (January 1st to December 31st, 2018) for all sixty-five monitoring stations located across Malaysia's fourteen states obtained from the Malaysian government's Department of Environment (DOE), which is part of the Ministry of Environment. Table 1 displays the information about the monitoring stations, including the location and its category of location.

Table 1. Locations of Air Monitoring Stations in Malaysia

Location Station	Latitude	Longitude	Category
Batu Pahat	01° 55' 09.56" N	102° 51' 59.82" E	Sub Urban
Kluang	02° 02' 16.37" N	103° 18' 43.42" E	Urban
Kota Tinggi	01° 33' 50.60" N	104° 13' 31.10" E	Urban
Larkin	01° 29' 40.65" N	103° 44' 09.50" E	Industrial
Pasir Gudang	01° 28' 12.43" N	103° 53' 36.44" E	Industrial
Pengerang	01° 23' 22.16" N	104° 08' 58.50" E	Sub Urban
Segamat	02° 29' 38.09" N	102° 51' 45.69" E	Urban
Tangkak	02° 17' 41.26" N	102° 34' 17.74" E	Urban
Alor Setar	06° 08' 13.49" N	100° 20' 48.71" E	Urban
Kulim Hi-tech	05° 24' 05.82" N	100° 35' 22.70" E	Sub Urban
Langkawi	06° 19' 53.54" N	099° 51' 30.45" E	Sub Urban
Sungai Petani	05° 37' 46.63" N	100° 28' 03.83" E	Sub Urban
Kota Bharu	06° 08' 50.75" N	102° 14' 57.24" E	Urban
Tanah Merah	05° 48' 40.21" N	102° 08' 04.20" E	Industrial
Alor Gajah	02° 22' 15.33" N	102° 13' 28.53" E	Sub Urban
Bandaraya Melaka	02° 11' 27.36" N	102° 15' 25.40" E	Urban
Bukit Rambai	02° 16' 06.57" N	102° 11' 37.19" E	Industrial
Balok Baru	03° 57' 38.31" N	103° 22' 55.76" E	Industrial
Indera Mahkota	03° 49' 09.18" N	103° 17' 47.57" E	Sub Urban
Jerantut	03° 56' 54.09" N	102° 21' 59.87" E	Background
Rompin	02° 55' 35.92" N	103° 25' 09.11" E	Urban
Temerloh	03° 28' 17.77" N	102° 22' 35.06" E	Urban

Location Station	Latitude	Longitude	Category
Balik Pulau	05° 20' 13.61" N	100° 12' 59.21" E	Urban
Minden (USM)	05° 21' 22.35" N	100° 18' 28.51" E	Urban
Seberang Jaya	05° 23' 53.41" N	100° 24' 14.20" E	Sub Urban
Seberang Perai	05° 19' 45.68" N	100° 26' 36.51" E	Industrial
Pegoh	04° 33' 12.00" N	101° 04' 48.84" E	Urban
Seri Manjung	04° 12' 01.23" N	100° 39' 48.08" E	Sub Urban
Tanjung Malim	03° 41' 15.92" N	101° 31' 28.17" E	Sub Urban
Taiping	04° 53' 55.86" N	100° 40' 44.78" E	Industrial
Tasek Ipoh	04° 37' 45.99" N	101° 06' 59.94" E	Industrial
Kangar	06° 25' 47.71" N	100° 12' 39.84" E	Sub Urban
Keningau	05° 20' 21.54" N	116° 09' 49.16" E	Urban
Kota Kinabalu	05° 52' 46.87" N	116° 03' 23.13" E	Urban
Kimanis	05° 32' 17.60" N	115° 51' 02.00" E	Sub Urban
Sandakan	05° 51' 52.08" N	118° 05' 27.92" E	Sub Urban
Tawau	04° 14' 59.22" N	117° 56' 09.11" E	Urban
Bintulu	03° 10' 37.50" N	113° 02' 27.92" E	Sub Urban
ILP Miri	04° 29' 41.24" N	114° 02' 36.29" E	Rural
Kuching	01° 33' 44.02" N	110° 23' 20.24" E	Industrial
Kapit	02° 00' 52.19" N	112° 55' 38.49" E	Rural
Limbang	04° 45' 32.00" N	115° 00' 49.20" E	Sub Urban
Mukah	02° 52' 59.65" N	112° 01' 11.07" E	Sub Urban
Miri	04° 25' 28.84" N	114° 00' 44.73" E	Sub Urban
Sibu	02° 18' 51.86" N	111° 49' 54.89" E	Urban
Samalaju	03° 32' 13.41" N	113° 17' 42.60" E	Urban
Sarikei	02° 07' 58.11" N	111° 31' 22.33" E	Urban
Samarahan	01° 27' 17.47" N	110° 29' 29.41" E	Rural
Sri Aman	01° 13' 10.76" N	111° 27' 53.25" E	Urban
Banting	02° 49' 00.08" N	101° 37' 23.36" E	Sub Urban
Klang	03° 00' 53.60" N	101° 24' 47.19" E	Urban
Kuala Selangor	03° 19' 16.70" N	101° 15' 22.47" E	Sub Urban
Petaling Jaya	03° 07' 59.40" N	101° 36' 28.83" E	Industrial
Shah Alam	03° 06' 16.98" N	101° 33' 22.39" E	Sub Urban
Nilai	02° 49' 18.09" N	101° 48' 41.34" E	Industrial
Port Dickson	02° 26' 28.97" N	101° 52' 00.68" E	Urban
Seremban	02° 43' 24.17" N	101° 58' 06.58" E	Urban
Besut	05° 44' 54.41" N	102° 30' 56.27" E	Urban
Kuala Terengganu	05° 18' 29.13" N	103° 07' 13.41" E	Urban
Kemaman	04° 15' 43.46" N	103° 25' 32.90" E	Industrial
Paka	04° 35' 53.03" N	103° 26' 05.34" E	Industrial
Batu Muda	03° 12' 44.78" N	101° 40' 56.02" E	Urban

Location Station	Latitude	Longitude	Category
Cheras	03° 06' 22.44" N	101° 43' 04.50" E	Urban
Labuan	05° 19' 57.81" N	115° 14' 17.62" E	Urban
Putrajaya	02° 54' 53.33" N	101° 41' 24.17" E	Urban

Methodology

Time series can be regarded as the realization of a stochastic process that is, a series of random variables ordered in time. Many problems related to atmospheric systems among which PM_{10} is included deal with temporal data that need to be analyzed by means of time series analysis. The statistical models to describe and forecast PM_{10} data are given in this section and the framework of the methodology is depicted in Figure 1.

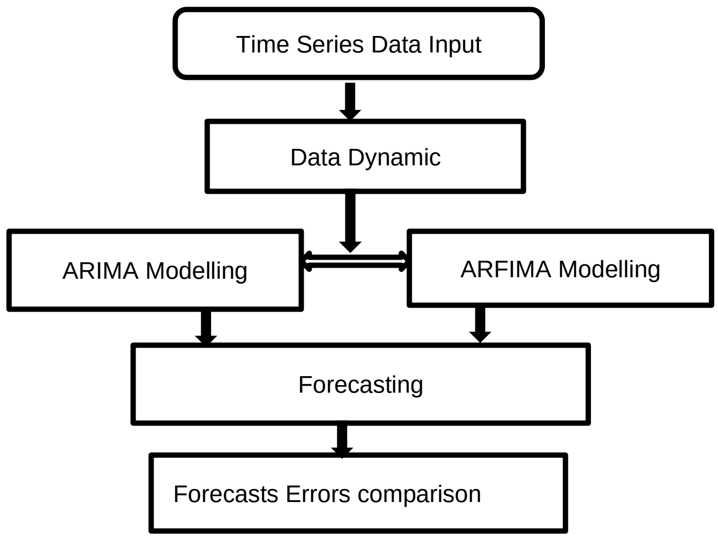


Figure 1. Flowchart summarizing the methodological framework.

Mann-Kendall Test

This test uses a time series data to verify an increase or decrease in time trend for the dataset. The test can therefore be used to investigate whether concentrations of contaminant is increasing or decreasing over time. The test compares the relative magnitude of the portion of dataset (sample) instead of the entire values of a given dataset (population).

The Mann-Kendall test statistic is defined as:

$$S_{mk} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} sign(x_j - x_i) \quad (1)$$

where

$$sign(x_j - x_i) = \begin{cases} 1, if\ (x_j - x_i) > 0 \\ 0, if\ (x_j - x_i) = 0 \\ -1, if\ (x_j - x_i) < 0 \end{cases}$$

and is the number of positive differences less the number of negative differences. If the sign $S = 0$ translates to mean that the trend of the dataset is neither increasing nor decreasing. The trend is decreasing if $S < 0$ indicating the decrease in the emission of pollutant over time interval. The emission of the contaminant is increasing over a given time interval when $S > 0$. The Mann-Kendall trend test is a distribution free and less sensitive to outliers than corresponding parametric approaches. The magnitude of the trend in the test is described by the Mann-Kendall τ estimator at 95% confidence level [28].

The Mann-Kendall τ statistic give the magnitude of the trend of a given series while the Mann-Kendall S statistic as defined in the equation (1) above can be used to tell whether the trend is increasing or decreasing. The trend is said to be increasing if the S value is positive and decreasing if it is negative.The probability value (p-value) of the test can be used to take appropriate decision about the rejecting or accepting the null hypothesis of no trend in the series. Rejecting the null hypothesis indicates the acceptance of an alternative hypothesis that the trend exists in the series at given level of significance.

Time Series Models

A time series process X at a time t (X_t), that is, $\{X_{t,}\, t = 0, \pm 1, \pm 2, \ldots\}$ is said to be stationary if the mean and the variance are time independent and the covariance between any two given observations depends on the temporal distance between them, not on their exact location in time. These requirements must be satisfied for any series to make statistical inference [30]. Given a zero-mean covariance stationary process $\{X_{t,}\, t = 0, \pm 1, \ldots\}$ with autocovariance function $\gamma_\mu = E[X_t,\, X_{t+\mu}]$, X_t is said to be integrated of order 0 if

$$\sum_{\mu=-\infty}^{\infty} |\gamma_\mu| < \infty. \tag{2}$$

In such a case, we denote $X_t \approx I(0)$. If a time series is nonstationary, a possibility of transforming the series to a stationary is to take its first differencing as

$$(1 - B)X_t = Y_t,$$

where B is the lag operator $(BX_t = X_{t-1},)$ and Y_t is $I(0)$. If the nonstationary series becomes stationary after first differencing, then X_t is said to be integrated of order 1 and is denoted as $X_t \approx I(1)$. Otherwise, more differencing may be needed to make the series stationary. If the series become stationary after the second differencing, we denote $X_t \approx I(2)$.

In some cases, the number of differencing required to make the series stationary is fractional, not an integer. In such a case, the process is said to be fractionally integrated, and X_t is $I(d)$.
That is,

$$(1 - B)^d X_t = Y_t\,. \tag{3}$$

Equation (3) can be expressed in terms of binomial expansion, such that $\forall d \in \mathbb{R}$. By the properties of binomial expansion,

$$(1 - B)^d = \sum_{r=0}^{d} \binom{d}{r} B^r (-1)^r$$

$$= 1 - dB + \frac{d(d-1)}{2} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \cdots$$

where B is lag operator, that is $BX_t = X_{t-1,}\, B^2 X_t = X_{t-2}, \ldots$.

Expanding equation (3) in terms of X_t can now be seen as

$$X_t = dX_{t-1} - \frac{d(d-1)}{2} X_{t-2} + \frac{d(d-1)(d-2)}{3!} X_{t-3} \ldots + Y_t\,. \tag{4}$$

If d is a positive integer value, X_t will be a function a finite number of past records, while if d is not an integer, X_t depends strongly on the values of time series data in previous far [31, 32]. However, the higher the value of d, the more the observations are going to be related between themselves [33]. To testify the stationarity condition, the Augmented Dickey-Fuller Test (ADF) was carried out for all series.

Estimation of Order of Integration

The procedure for estimating the degree of persistence in time series data developed by John Geweke and Susan Porter-Hudak [34] is employed for the purpose of this article. The test is semiparametric which is based on the simple linear regression of the log periodogram on a deterministic regressor and was found efficient by numerous researchers to estimate the degree of persistence [35, 30].

The spectral density function $f(\theta)$ of the fractionally integrated process $\{Y_t\}$ is given by,

$$f(\theta) = [2sin(\theta/2)]^{-2d} f_u(\theta), \quad (5)$$

where θ is the Fourier frequency, $f_u(\theta)$ is the spectral density corresponding to u_t, u_t is a stationary short memory noise with 0 mean.

Consider the set of harmonic frequencies $\theta_k = (2\pi_k/n)$, where $k = 0, 1, \ldots n/2$, n is the sample size. Applying the logarithm to both sides of equation (5) we obtain

$$\ln f(\theta_i) = \ln f_u(0) - d\ln[4sin^2(\theta_k/2)]. \quad (6)$$

Equation (5) can be seen as

$$\ln f(\theta_k) = \ln f_u(0) - d\ln[4sin^2(\theta_k/2)] + \ln\left[\frac{f_u(\theta_k)}{f_u(0)}\right]. \quad (7)$$

based on Wang *et al*. [36].

The parameter *d* for fractional differencing can be estimated by the regression equation as in (6) above. Using the periodogram estimate of $f(\theta_k)$, if the number of frequencies *p* used in equation (6) is a function $g(n)$, *n* sample size, $p = g(n) = n^x$ with $0 < x < 1$. The least squares estimate $\hat{d}$ using the above regression equation is asymptotically normally distributed in large sample and reliable in at least 50 observations [34]. That is,

$\hat{d} \sim N\left(d, \frac{\pi^2}{6\sum_{k=1}^{g(n)}(U_k - \bar{U})^2}\right)$, where $U_k = \ln[4sin^2(\theta_k/2)]$ and $\bar{U}$ is the sample U_k, $k = 1, \ldots g(n)$.

With the null hypothesis of short memory $(d = 0)$, the t statistic

$t_{d=0} = \hat{d}\left(\frac{\pi^2}{6\sum_{k=1}^{g(n)}(U_k - \bar{U})^2}\right)$ has a limiting standard normal distribution.

Autoregressive Moving Average (ARMA) Process

It is an important class of linear time series process that provides a general framework for studying stationary process. A stationary time series process is also known as ARMA process [37]. The process is said to be autoregressive integrated moving average (ARIMA) when it is non-stationary, and the series is differenced to make it stationary before modelling it with ARMA model. For dealing with univariate time series, the traditional ARMA algorithm is seen to be an efficient and reliable method. The ARMA process has advantage of not requiring any additional variables because it is based on the values of its historical observations, provided the series is stationary and the minimum number of observations in the data is at least fifty [38]. The process $\{X_t\}$ is said to be ARMA (p, q) if it is stationary and for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (8)$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, $(1 - \phi_1 Z - \cdots - \phi_p Z^p)$ and $(1 + \theta_1 Z + \cdots + \theta_q Z^q)$ are polynomials with no common factors [37].

Autoregressive Fractionally Integrated Moving Average (ARFIMA) Process

The ARFIMA (*p, d, q*) process is one of the best-known classes of long-memory models. The process is commonly employed in long-range dependence (LRD) time series modeling, particularly for high frequency air pollution data, network traffic, and hydrology datasets, among other things. In practice, several time series exhibit LRD in their data, prompting the development of several estimation and prediction approaches to account for the slowly decaying autocorrelations [39].
An ARFIMA process $\{y_t\}$ can be defined as:

$$\phi(B)y_t = \theta(B)(1 - B)^{-d}\varepsilon_t. \quad (9)$$

Here, $\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ are the autoregressive and moving average polynomials respectively, while $(1-B)^{-d}$ is an operator representing fractional that can be defined by the binomial expansion as

$$(1-B)^{-d} = \sum_{j=0}^{\infty} \eta_j B^j = \eta(B),$$

where $\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(1)}$, $d < \frac{1}{2}$, and $\{\varepsilon_t\}$ is error term with stationary (finite) variance.

Long memory is feature of time series that are characterized by a high degree of dependence between observations at higher lags. It comprehensively covers different statistical issues of time series such as short memory $(d = 0)$. If $-0.5 < d < 0.5$, the process is stationary and ergodic with a bounded and positively valued spectrum at all frequencies. The process has a short memory, and the sum of absolute autocorrelations is constant when $-0.5 < d < 0$. For $0 < d < 0.5$, the process exhibit long memory in the sense of equation (3) since its autocorrelations are all positive and decay at hyperbolic rate. For $0.5 \leq d < 1$, the process is non-stationary and mean reverting, unit roots $(d = 1)$, and has explosive or permanent effects when $d \geq 1$ [30, 40].

Results and Discussion

Descriptive Statistics

Table 2 provides the descriptive statistics of PM_{10} emission from monitoring stations across Malaysian states. In each of the stations, average concentration was calculated to observe station with highest concentration of the pollutant. Minimum and maximum values for each of the stations were also identified, lastly the standard deviation was also calculated to measure the level of disparity between the daily records from their respective mean values. Taiping and Tanjung Malim monitoring stations produced highest and least average concentration respectively. Tanjung Malim station has the least volatile series while Klang station has the most volatile series of PM_{10} emission during the period and locations covered by the study.

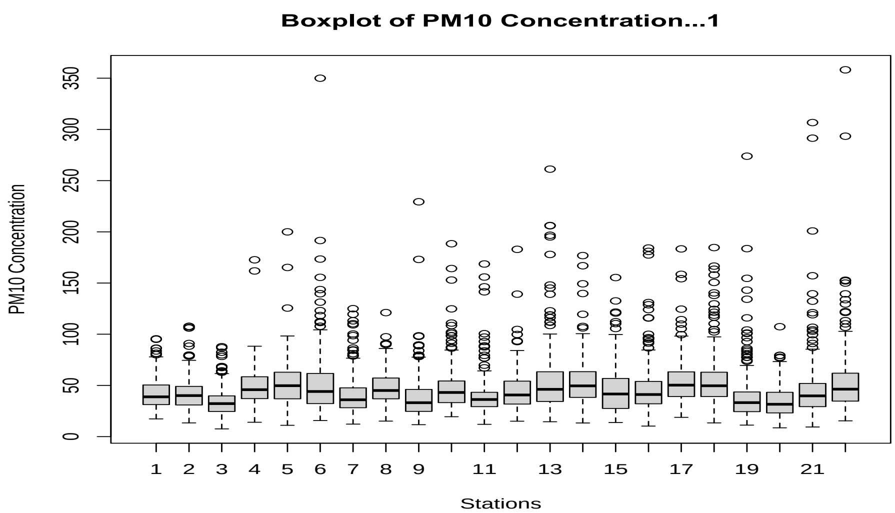
Table 2. Descriptive statistics of PM_{10} pollutant for sixty-five Malaysian monitoring stations in daily scale for the period January to December 2018

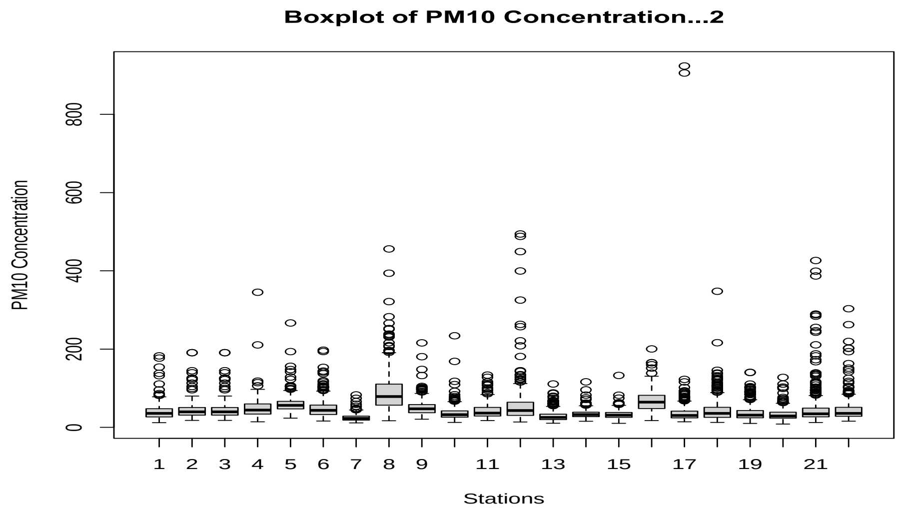
Station	Mean	Minimum	Maximum	Std. Dev	Category
Batu Pahat	41.49	17.33	95.49	13.71	Sub Urban
Kluang	41.73	13.47	107.74	15.08	Urban
Kota Tinggi	33.68	7.57	87.82	12.75	Urban
Larkin	48.82	14.09	172.66	16.85	Industrial
Pasir Gudang	51.18	11.08	199.95	19.69	Industrial
Pengerang	50.64	15.66	349.96	29.37	Sub Urban
Segamat	39.88	12.22	125.00	17.43	Urban
Tangkak	49.26	15.17	122.21	16.89	Urban
Alor Setar	37.76	11.70	229.31	20.32	Urban
Kulim Hi-Tech	46.56	19.45	188.36	20.12	Sub Urban
Langkawi	38.52	12.06	168.60	17.82	Sub Urban
Sungai Petani	44.15	15.08	182.92	17.99	Sub Urban
Kota Bharu	53.67	14.60	261.26	30.60	Urban
Tanah Merah	52.82	13.32	176.82	22.31	Industrial
Alor Gajah	45.02	13.82	208.60	23.45	Sub Urban

Station	Mean	Minimum	Maximum	Std. Dev	Category
Bandaraya Melaka	45.65	10.33	184.19	22.10	Urban
Bukit Rambai	53.45	18.85	183.30	20.57	Industrial
Balok Baru	54.57	13.46	184.61	24.70	Industrial
I. Mahkota	37.42	11.26	273.87	24.20	Sub Urban
Jerantut	34.19	8.62	107.33	14.62	Background
Rompin	44.58	9.42	306.66	28.58	Urban
Temerloh	52.95	15.40	358.17	31.12	Urban
Balik Pulau	39.79	12.01	182.61	20.54	Urban
Minden (USM)	45.03	17.76	190.96	22.07	Urban
Seberang Jaya	49.37	14.41	239.36	20.80	Sub Urban
Seberang Perai	49.32	14.12	345.30	25.76	Industrial
Pegoh	59.52	23.56	266.73	22.29	Urban
Seri Manjung	49.40	16.29	197.18	25.10	Sub Urban
Tanjung M.	24.44	11.46	82.81	9.02	Sub Urban
Taiping	92.37	16.83	456.03	54.44	Industrial
Tasek Ipoh	50.33	20.73	215.99	20.82	Industrial
Kangar	36.54	12.61	233.97	19.13	Sub Urban
Keningau	42.21	17.32	133.60	18.84	Urban
Kinabalu	28.94	10.54	110.83	13.80	Urban
Kimanis	28.94	10.54	110.83	13.80	Sub Urban
Sandakan	34.34	15.50	116.18	11.11	Sub Urban
Tawau	32.56	10.24	132.71	11.55	Urban
Bintulu	67.42	17.29	200.31	27.61	Sub Urban
ILP Miri	40.33	14.24	923.54	67.21	Rural
Kuching	44.72	12.68	347.70	31.73	Industrial
Kapit	36.87	10.16	140.58	19.82	Rural
Limbang	36.87	10.16	140.58	19.82	Sub Urban
Mukah	49.75	12.42	426.51	51.50	Sub Urban
Miri	45.48	15.68	303.22	32.95	Sub Urban
Sibu	49.93	17.12	718.93	45.87	Urban
Samalaju	53.24	19.00	293.93	31.76	Urban
Sarikei	40.44	13.89	386.61	32.53	Urban
Samarahan	36.63	14.67	141.26	18.99	Rural
Sri Aman	35.44	11.02	148.02	19.39	Urban
Banting	60.91	19.22	173.82	23.40	Sub Urban
Klang	90.64	29.29	1453.98	126.02	Urban
Kuala	46.95	11.57	330.67	23.38	Sub Urban
Petaling Jaya	57.49	22.03	130.95	16.99	Industrial
Shah Alam	57.70	22.87	168.49	16.62	Sub Urban

Station	Mean	Minimum	Maximum	Std. Dev	Category
Nilai	73.59	28.18	319.87	29.79	Industrial
Port Dickson	39.53	17.49	117.45	13.66	Urban
Seremban	42.65	13.83	117.70	16.73	Urban
Besut	42.30	11.57	255.19	22.17	Urban
Kuala	51.92	12.22	905.76	50.87	Urban
Kemaman	46.68	14.49	488.44	36.45	Industrial
Paka	37.15	13.18	120.21	15.80	Industrial
Batu Muda	49.97	14.18	129.14	19.01	Urban
Cheras	57.15	22.53	291.82	21.51	Urban
Labuan	45.86	11.68	257.96	30.31	Urban
Putrajaya	53.15	21.13	221.06	22.92	Urban

Figure 2 give the PM_{10} representation on boxplot for all the sixty-five monitoring stations. The lower and upper outlier values for each of the monitoring stations can be seen from the boxplot graphical representation. Minimum and maximum emission were also displayed alongside the first, second and third quartile deviations. It can be seen from the graph that significant number of PM_{10} emissions exceeded the World Health Organization (WHO) standard of $50\mu g/m^3$.





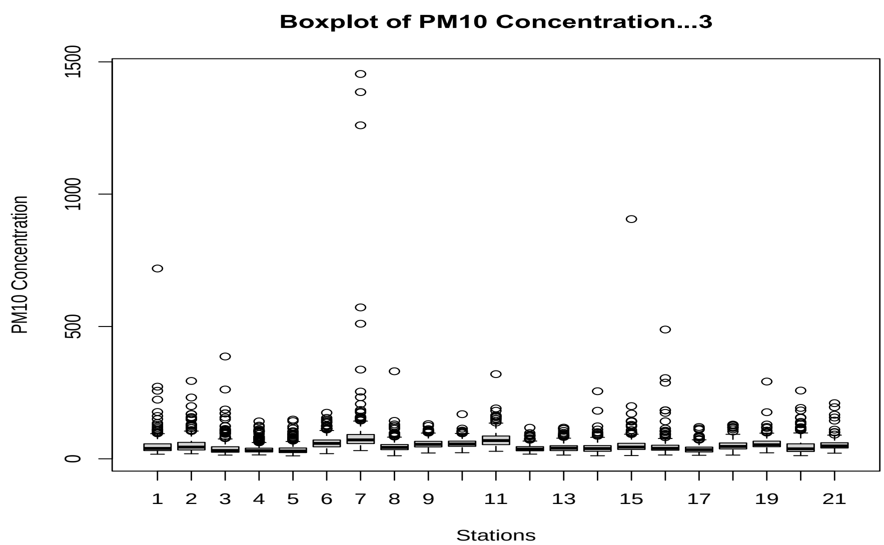


Figure 2. Boxplot summarizing PM_{10} concentration in sixty-five monitoring stations

Figure 2 successfully presents the pictorial distribution of PM_{10} emission in all the sixty-five monitoring stations including their outlier values. It also gives for each monitoring station; first quartile, second quartile (median), and the third quartile. Generally, the concentration of outlier values in all the considered monitoring stations is from the right direction, that is above the upper fence of the boxplot graphical representation of each monitoring station.

The results of long memory and Mann-Kendall trend analyses of the PM_{10} emission are presented in Table 3. The degree of persistence of PM_{10} pollutant across Malaysia was estimated using parameter *d* of long memory analyses at 95% confidence bands. The estimated results of persistence are in the interval $(0,1)$ which indicates evidence of long memory and fractional integration. The result of long memory analysis reveals the evidence of persistence in the emission of the pollutant in most of the monitoring sites at $(0 < d < 0.5)$. This is preceded by number of stations with non-stationary process but mean reverting at $(0.5 \leq d < 1)$ which translate to mean that the adverse effects (shocks) of pollutant is transitory not permanent, while Keningau station in the state of Sabah has a stationary short memory process. The shocks in the series remain transitory so long the parameter $d < 1$, but permanent or explosive when $d \geq 1$. However, the higher the value of *d* the higher the degree of persistence and vice versa [41]. In the overall, the result of our analysis reveals strong evidence of persistence in the level of PM_{10} emission but with transitory (not permanent) shocks in sixty-four of the sixty-five monitoring sites across Malaysia. This means strong anti-pollution measures need to be strengthened to combat the persistence though mean reverting effects of the PM_{10} pollutant.

The results of Mann-Kendall trend test on the other hand, reveals the evidence of increasing and decreasing trend in the emission of PM_{10} pollutant at most of the monitoring sites, with evidence of no trend in twenty-four of the sixty-five monitoring sites representing 37%. The monotonic trend increases in seventeen monitoring sites (Kluang, Kota Tinggi, Larkin, Pasir Gudang, Alor Gajah, Temerloh, Seberang Jaya, Seberang Perai, Kota Kinabalu, ILP Miri, Kuching, Kapit, Mukah, Miri, Sri Aman, Nilai and Batu Muda) representing 26% and decreases in remaining twenty-four monitoring sites representing 37%. Decreasing trend in the emission of PM_{10} pollutant translates to mean the corresponding decrease in the adverse effects of pollutant on human health, ambient air quality and the environment. It is therefore recommended for the government to maintain the anti-pollutant measures put in place at sites with evidence of decreasing trend and strengthen those measures at the sites of other two cases.

Table 3. The results of Long Memory and Mann – Kendall trend analyses of PM_{10} pollutant for sixty-five Malaysian monitoring stations in daily scale during January to December 2018

Station	*d* value	Kendall's τ	Kendall's *S*	*p*-value	Trend Interpretation
Batu Pahat	0.450	0.0333	2212	0.343	No Trend
Kluang	0.358	0.0745	4949	0.034	Increasing Trend
Kota Tinggi	0.310	0.1529	10159	<0.0001	Increasing Trend
Larkin	0.859	0.2810	18667	<0.0001	Increasing Trend
Pasir Gudang	0.667	0.2028	13470	<0.0001	Increasing Trend
Pengerang	0.424	0.0305	2026	0.385	No Trend
Segamat	0.112	0.0512	3403	0.144	No Trend
Tangkak	0.153	-0.0061	-0.0061	0.877	No Trend
Alor Setar	0.199	-0.1611	-10705	<0.0001	Decreasing Trend
Kulim Hi-Tech	0.287	-0.1370	-9099	<0.0001	Decreasing Trend
Langkawi	0.331	-0.2437	-16190	<0.0001	Decreasing Trend
Sungai Petani	0.086	0.0593	3938	0.091	No Trend
Kota Bharu	0.606	-0.1115	-7407	0.002	Decreasing Trend
Tanah Merah	0.210	-0.1305	-8670	0.0002	Decreasing Trend
Alor Gajah	0.490	0.1215	8071	0.001	Increasing Trend
Bandaraya Melaka	0.406	-0.1220	-8102	0.001	Decreasing Trend
Bukit Rambai	0.363	0.0626	4156	0.074	No Trend
Balok Baru	0.622	-0.1822	-12106	<0.0001	Decreasing Trend
Indera Mahkota	0.432	-0.0812	-5394	0.021	Decreasing Trend
Jerantut	0.347	0.0148	981	0.674	No Trend
Rompin	0.502	-0.0195	-1294	0.579	No Trend
Temerloh	0.094	0.0685	4553	0.051	Increasing Trend
Balik Pulau	0.342	-0.0583	-3871	0.097	Decreasing Trend
Minden (USM)	0.286	-0.1214	-5700	0.002	Decreasing Trend
Seberang Jaya	0.457	0.1893	12578	<0.0001	Increasing Trend

Station	*d* value	Kendall's τ	Kendall's *S*	*p*-value	Trend Interpretation
Seberang Perai	0.673	0.0680	4516	0.053	Increasing Trend
Pegoh	0.547	-0.0271	-1798	0.44	No Trend
Seri Manjung	0.260	-0.1198	-7960	0.001	Decreasing Trend
Tanjung Malim	0.261	-0.1868	-12406	<0.0001	Decreasing Trend
Taiping	0.250	-0.1424	-9459	<0.0001	Decreasing Trend
Tasek Ipoh	0.328	-0.1390	-9233	0.0001	Decreasing Trend
Kangar	0.117	-0.1468	-9753	<0.0001	Decreasing Trend
Keningau	-0.031	-0.0292	-1938	0.406	No Trend
Kota Kinabalu	0.261	0.1702	11304	<0.0001	Increasing Trend
Kimanis	0.132	0.0496	3293	0.158	No Trend
Sandakan	0.211	-0.0433	-2875	0.217	No Trend
Tawau	0.632	-0.0846	-5623	0.016	Decreasing Trend
Bintulu	0.253	0.0447	2968	0.203	No Trend
ILP Miri	0.345	0.2150	14283	<0.0001	Increasing Trend
Kuching	0.535	0.1230	8169	0.001	Increasing Trend
Kapit	0.312	0.1390	9234	0.0001	Increasing Trend
Limbang	0.473	0.0518	3438	0.14	No Trend
Mukah	0.440	0.1857	12333	<0.0001	Increasing Trend
Miri	0.480	0.1841	12227	<0.0001	Increasing Trend
Sibu	0.383	0.0490	3258	0.162	No Trend
Samalaju	0.399	-0.0007	-49	0.984	No Trend
Sarikei	0.430	-0.0123	-814	0.727	No Trend
Samarahan	0.491	0.0575	3818	0.101	No Trend
Sri Aman	0.546	0.2047	13595	<0.0001	Increasing Trend
Banting	0.450	0.0130	864	0.711	No Trend
Klang	0.471	-0.1733	-11509	<0.0001	Decreasing Trend
Kuala Selangor	0.399	-0.0291	-1936	0.406	No Trend
Petaling Jaya	0.296	-0.1622	-10775	<0.0001	Decreasing Trend
Shah Alam	0.152	-0.0762	-5061	0.030	Decreasing Trend
Nilai	0.657	0.2147	14261	<0.0001	Increasing Trend
Port Dickson	0.432	-0.0172	-1142	0.624	Decreasing Trend
Seremban	0.282	0.0178	1184	0.612	No Trend
Besut	0.206	-0.1045	-6944	0.003	Decreasing Trend
Kuala Terengganu	0.004	-0.1232	-8186	0.0004	Decreasing Trend
Kemaman	0.228	-0.0464	-3080	0.186	No Trend
Paka	0.407	-0.0108	-715	0.759	No Trend

Station	*d* value	Kendall’s τ	Kendall’s *S*	*p*-value	Trend Interpretation
Batu Muda	0.313	0.1587	10543	<0.0001	Increasing Trend
Cheras	0.524	-0.1461	-9708	<0.0001	Decreasing Trend
Labuan	0.132	-0.0820	-3853	0.032	Decreasing Trend
Putrajaya	0.072	0.0206	1367	0.558	No Trend

In Table 3, the *p*-value in each case of monitoring station was compared with 5% level of significance. Null hypothesis of no trend in the emission of pollutant is rejected when the *p*-value is less than 0.05 level of significance concluding that there is evidence of increase or decrease in the trend of PM_{10} at given monitoring station depending on the corresponding value of S statistic. On the other hand, the null hypothesis is not rejected if the *p*-value is greater than 0.05 level of significance concluding that the trend of the series of PM_{10} is neither increasing nor decreasing.

ARIMA and ARFIMA Modelling

The dataset for ARIMA model needs to be stationary which means realization of constant mean and variance. The results from stationarity test by the ADF for some of the monitoring stations are presented in Table 4. Thus, the non-stationary series are transformed to stationary by means of differencing. The differencing is whether fractional or once to produce a stationary data.

Table 4. Results of ADF Test

Station	Statistic	Lag	P-value	Status
Shah Alam	-3.3085	36	0.06982	Non-stationary
Cheras	-1.9754	36	0.5873	Non-stationary
Tanjung Malim	-1.6865	36	0.7091	Non-stationary
Kota Tinggi	-0.9184	36	0.9505	Non-stationary
Kuching	-2.5152	36	0.3595	Non-stationary
Alor Gajah	-0.5869	36	0.9776	Non-stationary
Pasir Gudang	-0.747	36	0.9658	Non-stationary
Pegoh	-2.0375	36	0.561	Non-stationary
Nilai	-1.6521	36	0.7237	Non-stationary
Taiping	-3.269	36	0.0765	Non-stationary
Tawau	-1.0136	36	0.9356	Non-stationary
Kulim Hi-Tech	-2.7736	36	0.2504	Non-stationary

In each monitoring station, 358 observations were used to develop ARIMA and ARFIMA models, while the last seven observations were used to assess the performance of the developed model using forecast error analysis, which involved comparing the forecasted with the actual observations. The three measures of error accuracy used for this research for assessing forecasting are Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Due to the presence of outliers as observed in most PM_{10} dataset across the monitoring stations, RMSE should be regarded as the best measure of forecast accuracy over others [42, 43]. Any of the two models with the least RMSE value in the column is then identified as the best performing model.

Table 5 displays the ARIMA and ARFIMA model parameters, Ljung-Box test result to confirm models' adequacy and three measures of error accuracy to determine the best performing model in each monitoring station separately. The fractional integration parameter *d* values for each ARFIMA model were also indicated in model structure. The two models were observed to have adequately fitted the PM_{10} concentration from the sixty-five monitoring stations across Malaysia due to achieving uncorrelated model residuals using Ljung-Box test statistic at 5% level of significance. We failed to reject the null hypothesis of uncorrelated model residuals in all the developed ARIMA and ARFIMA models.

Table 5. Seven-day ahead Forecast and Results of Error Analysis

Station	**ARIMA/ ARFIMA**	**Ljung Box (*p*-value)**	**MAD**	**MAPE (%)**	**RMSE**
Segamat	1, 1, 1	0.8104	10.00	24.76	12.84
	2, 0.291, 1	0.8236	9.62	27.08	11.71
Batu Pahat	2, 1, 1	0.9440	11.38	26.24	12.25
	1, 0.280, 1	0.1969	9.37	24.47	10.25
Kluang	2, 0, 1	0.9987	18.03	40.31	19.47
	1, 0.219, 2	0.6262	17.00	38.94	18.94
Larkin	2, 1, 1	0.9570	11.12	28.53	14.17
	2, 0.354, 1	0.9194	11.74	30.26	14.51
Pasir Gudang	2, 1, 2	0.9865	6.07	13.64	7.63
	3, 0.363, 0	0.1605	7.64	18.53	9.23
Pengerang	1, 1, 1	0.9973	9.90	26.23	11.70
	2, 0.329, 1	0.9827	9.39	28.79	11.28
Kota Tinggi	2, 1, 1	0.8865	5.36	19.65	6.27
	1, 0.351, 2	0.9195	6.36	18.67	7.07
Tangkak	3, 0, 0	0.9349	27.39	30.95	32.81
	3, 0.251, 0	0.4138	29.14	32.97	34.52
Langkawi	0, 1, 3	0.8237	6.18	17	8.35
	1, 0.270, 3	0.8571	6.42	20.63	7.61
Alor Setar	1, 0, 2	0.8815	7.86	27	9.67
	1, 0.302, 2	0.6413	6.66	22.37	7.49
Sungai Petani	1, 0, 2	0.9030	17.59	33	24.57
	2, 0.33, 1	0.4423	17.71	33.30	24.65
Kulim Hi-Tech	3, 1, 1	0.9888	7.61	18.57	8.71
	3, 0.116, 1	0.9794	6.32	16.31	8.27
Tanah Merah	1, 1, 1	0.2714	15.36	25.21	21.17
	3, 0.409, 3	0.4984	15.76	30.99	19.16
Kota Bharu	2, 1, 1	0.9279	22.81	25.87	32.03
	2, 0.145, 1	0.3927	22.05	26.83	29.24

Station	**ARIMA/ ARFIMA**	**Ljung Box (*p*-value)**	**MAD**	**MAPE (%)**	**RMSE**
Alor Gajah	0, 1, 3	0.8115	31.06	28.81	66.76
	2, 0.341, 0	0.2183	28.50	27.16	63.87
Bukit Rambai	2, 1, 1	0.9905	8.92	22.37	10.75
	2, 0.289, 1	0.4541	8.36	23.61	11.83
Bandaraya Melaka	1, 1, 2	0.8188	6.04	21.54	7.32
	1, 0.366, 2	0.8120	6.79	24.87	8.35
Rompin	5, 1, 2	0.8763	16.29	51.44	17.83
	6, 0.423, 3	0.9597	16.55	53.16	17.62
Temerloh	1, 1, 2	0.9965	4.71	11.73	5.78
	1, 0.213, 2	0.7369	6.80	17.60	8.03
Jerantut	3, 1, 1	0.9972	6.98	30.05	8.59
	3, 0.4, 1	0.7544	7.97	34.49	9.42
Indera Mahkota	1, 1, 1	0.8113	3.09	14.35	3.37
	1, 0.256, 1	0.1701	7.36	34.33	7.86
Balok Baru	0, 1, 2	0.8814	9.56	35.22	12.77
	2, 0.354, 1	0.9675	13.46	46.26	15.81
Seberang Jaya	1, 1, 1	0.7895	13.63	33.57	17.72
	2, 0.358, 1	0.7746	12.84	32.12	17.42
Seberang Perai	1, 1, 1	0.7570	9.68	21.10	13.01
	3, 0.353, 1	0.2858	11.09	25.86	12.88
Minden (USM)	2, 0, 5	0.8805	5.95	22.24	9.53
	1, 0.373, 2	0.5809	5.56	21.41	8.73
Balik Pulau	1, 0, 2	0.9627	6.37	19.81	7.39
	1, 0.285, 2	0.8923	5.60	17.46	6.73
Pegoh	1, 0, 2	0.9268	8.78	19.02	10.63
	3, 0.306, 0	0.2766	9.06	19.79	10.57
Seri Manjung	1, 0, 1	0.9442	14.44	50.54	15.86
	2, 0.171, 1	0.6588	12.73	44.70	13.97
Taiping	0, 1, 3	0.9794	48.14	38.68	51.48
	2, 0.176, 1	0.8621	32.62	25.39	37.48
Tanjung Malim	2, 1, 1	0.9646	3.67	16.29	4.25
	2, 0.318, 0	0.0996	2.43	11.22	2.78
Tasek Ipoh	1, 1, 2	0.8729	5.82	12.28	7.47
	1, 0.339, 2	0.7704	5.56	12.75	7.09
Kangar	1, 0, 0	0.6686	3.66	8	6.13
	1, 0.003, 1	0.4722	3.81	8.47	6.03
Tawau	1, 1, 2	0.9846	5.99	29.58	7.23

Station	ARIMA/ ARFIMA	Ljung Box (*p*-value)	MAD	MAPE (%)	RMSE
	1, 0.379, 2	0.7276	7.65	40.86	8.60
Sandakan	1, 0, 0	0.8488	9.76	32.35	11.68
	2, 0.127, 1	0.9985	8.93	28.28	11.30
Kota Kinabalu	1, 1, 3	0.6899	25.47	67.15	30.81
	0, 0.057, 9	0.0969	22.17	54.68	29.48
Kimanis	1, 0, 1	0.9869	5.67	33.99	7.11
	2, 0.066, 2	0.1677	5.91	35.40	7.32
Keningau	1, 0, 2	0.9387	9.25	38.73	11.74
	1, 0.200, 2	0.6223	8.71	36.78	11.29
Limbang	1, 0, 1	0.4243	4.32	19.29	6.28
	1, 0.109, 1	0.0886	4.12	18.39	6.07
ILP Miri	0, 1, 2	0.7313	10.22	37.42	11.35
	1, 0.119, 0	0.1152	12.35	48.14	12.73
Miri	0, 1, 1	0.4147	23.69	58.52	33.33
	3, 0.264, 0	0.1204	23.31	75.42	33.28
Samalaju	3, 1, 2	0.9725	21.26	47.24	27.41
	3, 0.308, 2	0.9932	21.40	47.07	27.39
Bintulu	1, 0, 1	0.8890	18.31	53.26	23.66
	1, 0.153, 1	0.0594	17.67	51.33	22.80
Mukah	1, 1, 1	0.4673	11.80	48.38	12.56
	2, 0.279, 0	0.2483	16.74	68.06	17.54
Kapit	2, 1, 1	0.8827	9.14	22.42	12.33
	2, 0.126, 1	0.4428	9.57	24.94	11.95
Sibu	1, 1, 2	0.9965	9.73	28.71	10.28
	1, 0.287, 2	0.9770	10.04	29.89	10.66
Sarikei	1, 0, 2	0.9323	8.37	42.07	10.21
	1, 0.001, 2	0.9120	8.53	42.79	10.36
Sri Aman	1, 1, 2	0.7817	12.33	53.89	12.90
	1, 0.174, 2	0.8483	13.30	57.91	13.78
Samarahan	1, 1, 2	0.9156	4.89	18.41	6.26
	1, 0.120, 2	0.9283	6.39	25.28	7.64
Kuching	0, 1, 2	0.9886	13.25	27.62	18.06
	2, 0.191, 1	0.9241	12.53	29.45	16.39
Kuala Selangor	1, 0, 2	0.9672	15.55	43.57	15.98
	2, 0.37, 0	0.0500	13.79	37.73	14.57
Petaling Jaya	1, 1, 1	0.9838	7.53	14.19	9.42
	2, 0.269, 1	0.9864	7.99	16.58	8.82

Station	ARIMA/ ARFIMA	Ljung Box (p-value)	MAD	MAPE (%)	RMSE
Shah Alam	2, 0, 2	0.9035	11.76	29.56	13.47
	2, 0.229, 2	0.2852	11.41	28.76	13.20
Klang	1, 0, 1	0.4682	29.28	71.30	31.35
	2, 0.330, 2	0.9848	29.17	71.07	31.27
Banting	2, 1, 1	0.9463	16.14	28.69	18.77
	2, 0.336, 0	0.0954	15.37	29.99	17.06
Nilai	0, 1, 2	0.9602	10.69	16.94	13.95
	2, 0.362, 2	0.1153	10.10	15.81	12.93
Seremban	1, 0, 2	0.9447	13.69	24.39	17.03
	3, 0.204, 0	0.2013	14.09	24.54	17.83
Port Dickson	1, 0, 2	0.9285	8.54	19.87	9.77
	2, 0.234, 2	0.9785	8.69	19.91	10.14
Kemaman	1, 0, 1	0.8645	12.67	39.35	13.69
	1, 0.191, 1	0.1963	10.06	30.86	11.12
Paka	3, 1, 1	0.7880	12.23	30.23	15.11
	3, 0.157, 1	0.7050	13.11	37.22	14.50
Kuala Terengganu	0, 1, 1	0.4190	27.14	36.25	32.84
	2, 0.117, 1	0.9828	23.77	33.33	29.29
Besut	1, 1, 2	0.9835	14.36	29.02	16.37
	1, 0.025, 1	0.1138	13.21	29.10	15.12
Batu Muda	2, 1, 1	0.6686	8.53	18.94	11.07
	2, 0.335, 1	0.8081	9.98	24.52	10.59
Cheras	1, 1, 1	0.9855	11.42	20.98	15.39
	1, 0.278, 1	0.8081	11.85	25.51	13.99
Putrajaya	2, 0, 2	0.9510	43.16	36.36	72.45
	2, 0.212, 1	0.8927	43.09	36.12	72.48
Labuan	1, 0, 1	0.6967	23.42	141.56	24.31
	2, 0.041, 1	0.8320	26.11	155.38	26.72

p-value < 0.05 means rejecting the null hypothesis of uncorrelated residual.

The forecasts result indicates that the ARFIMA model performed better than ARIMA model in 42 of the 65 modelled stations by producing least RMSE.

Conclusions

We analyzed concentration of PM_{10} air pollutant due to its adverse impact on human health, environment and ambient air quality using the dataset from monitoring stations across Malaysia for the year 2018. Boxplot graphical representation of the dataset were given in Figure 2 to visualize the statistical behavior of the dataset. Some descriptive statistics for each of the sixty-five monitoring stations were derived to display total number of observations used in each station, average concentration, minimum and maximum daily recorded emissions, and the standard deviation as contained in Table 2. The daily emission of the pollutant in most monitoring stations for some days were observed to have gone beyond the WHO standard of $50\mu g/m^3$.

The PM_{10} concentration for the reviewed period exhibited both LRD and SRD characteristics with dominating evidence of LRD over the SRD in most of the monitoring stations. The degree of persistence with orders of integration $(0,1)$ was significant in sixty-four of the sixty-five monitoring stations, this implies that shocks in the level of pollutant is not permanent but transitory. The fluctuation in the level of contaminant in daily scale is not statistically significant in twenty-four monitoring stations but significantly increasing and decreasing at seventeen and remaining twenty-four monitoring stations respectively as reported and confirmed by Mann-Kendall trend test at 5% level of significance. Each station was analyzed and forecasted using both ARIMA and ARFIMA models. The results revealed that ARFIMA model performed better in forty-two of the sixty-five stations. The result of the Ljung-Box test of uncorrelated residuals confirms the adequacy of each of the ARIMA and ARFIMA models used in this study at 5% level of significance. In addition, models' performances were observed to be independent of category, and location of the given monitoring stations. This conclude to mean that none of the models used for this research can be regarded as universal in forecasting and analyzing PM_{10} concentration in Malaysia.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

References

[1] Department of Environment, Malaysia (2018). *Laporan Kualiti Alam Sekeliling*. https://www.academia.edu/download/55174498/22232-63936-1-PB.pdf.

[2] Fong, S. Y., Abdullah, S., Ismail, M. (2018). Forecasting of particulate matter (PM_{10}) concentration based on gaseous pollutants and meteorological factors for different monsoons of urban coastal area in Terengganu. *Journal of Sustain. Sci. Manag., 5*, 3-17. https://www.researchgate.net/profile/Samsuri-Abdullah-2/publication/330999018.

[3] Abdullah, S., Ismail, M., Natasha, N., Samat, A., Ahmed, A. N. (2018). Modelling particulate matter (PM_{10}) concentration in industrialized area: A comparative study of linear and nonlinear algorithms. *Journal of Engineering and Applied Sciences, 13*(20), 8227-8235. https://www.researchgate.net/profile/Samsuri-Abdullah-2/publication/328781284.

[4] Tajudin, M. A. B. A., Khan, M. F., Mahiyuddin, W. R. W., Hod, R., Latif, M. T., Hamid, A. H., Rahman, S. A., Sahani, M. (2019). Risk of concentrations of major air pollutants on the prevalence of cardiovascular and respiratory diseases in urbanized area of Kuala Lumpur, Malaysia. *Ecotoxicology and Environmental Safety 171*, 290-300. https://doi.org/10.1016/j.ecoenv.2018.12.057.

[5] Usmani, R. S. A., Saeed, A., Abdullahi, A. M., Pillai, T. R., Jhanjhi, N. Z., Hashem, I. A. T. (2020). Air pollution and its Health Impacts in Malaysia: A Review. *Air Qual Atmos Health,* 13(9), 1093-1118. https://doi.org/10.1007/s11869-020-00867-x.

[6] Masseran, N., Mohd-Safari, M. A. (2020). Intensity–duration–frequency approach for risk assessment of air pollution events. *Journal of Environmental Management, 264*, 110429 https://doi.org/10.1016/j.jenvman.2020.110429.

[7] Fauziah, S. H., Rizman-Idid, M., Cheah, W., Loh, K. H., Sharma, S., Noor, N. M., Bordt, M., Praphotjanaporn, T., Samah, A. A., Sabaruddin, J. S., George, M. (2021). Marine debris in Malaysia: A review on the pollution intensity and mitigating measures. *Marine Pollution Bulletin, 167*, 112258. https://doi.org/10.1016/j.marpolbul.2021.112258.

[8] 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. https://doi.org/10.1016/j.atmosenv.2020.117343.

[9] Khamraev, K., Cheriyan, D., Choi, J. H. (2021). A Review on health risk assessment of PM in the construction industry – Current situation and future directions. *Science of the Total Environment, 758*, 143716. https://doi.org/10.1016/j.scitotenv.2020.143716.

[10] Lam, H. C. Y., Jarvis, D., Fuertes, E. (2021). Interactive effects of allergens and air pollution on respiratory health: A systematic review. *Science of the Total Environment, 757*, 143924. https://doi.org/10.1016/j.scitotenv.2020.143924.

[11] Soleimani, Z., Boloorani, A. D., Khalifeh, R., Teymouri, P., Mesdaghinia, A., Griffin, D. W. (2019). Air Pollution and respiratory hospital admissions in Shiraz, Iran, 2009 to 2015. *Atmospheric Environment, 209*, 233-239. https://doi.org/10.1016/j.atmosenv.2019.04.030.

[12] Xu, J., Geng, W., Geng, X., Cui, L. J., Ding, T., Xiao, C., Zhang, J., Tang, J., Zhai, J. (2020). Study on the association between ambient air pollution and daily cardiovascular death in Hefei, China. *Environmental Science and Pollution Research, 27*(1), 547-561. https://doi.org/10.1007/s11356-019-06867-4.

[13] Jahn, H. J., Schneider, A., Breitner, S., Eißner, R., Wendisch, M., Krämer, A. (2011). Particulate matter pollution in the megacities of the Pearl River Delta, China - A systematic literature review and health risk assessment. *International Journal of Hygiene and Environmental Health, 214*(4), 281-295. https://doi.org/10.1016/j.ijheh.2011.05.008.

[14] Zhang, L. W., Chen, X., Xue, X. D., Sun, M., Han, B., Li, C. P., Ma, J., Yu, H., Sun, Z. R., Zhao, L. J., Zhao, B. X., Liu, Y. M., Chen, J., Wang, P. P., Bai, Z. P., Tang, N. J. (2014). Long-term exposure to high particulate matter pollution and cardiovascular mortality: A 12-year cohort study in four cities in Northern China. *Environment International, 62*, 41-47. https://doi.org/10.1016/j.envint.2013.09.012.

[15] Fong, S. Y., Abdullah, S., Ismail, M. (2018). Forecasting of particulate matter (PM_{10}) concentration based on gaseous pollutants and meteorological factors for different monsoons of urban coastal area in Terengganu. *Journal of Sustain. Sci. Manag., 5*, 3-17. https://www.researchgate.net/profile/Samsuri-Abdullah-2/publication/330999018.

[16] Aryal, A., Harmon, A. C., Dugas, T. R. (2021). Particulate matter air pollutants and cardiovascular disease: strategies for intervention. *Pharmacology and Therapeutics, 223*, 107890. https://doi.org/10.1016/j.pharmthera.2021.107890.

[17] Li, H., Guo, B., Han, M., Tian, M., Zhang, J. (2015). Particulate matters pollution characteristic and the correlation between PM ($PM_{2.5}$, PM_{10}) and meteorological factors during the summer in Shijiazhuang. *Journal of Environmental Protection, 6*(5), 457-463. https://doi.org/10.4236/jep.2015.65044.

[18] Lv, L., Chen, Y., Han, Y., Cui, M., Wei, P., Zheng, M., Hu, J. (2021). High-time-resolution $PM_{2.5}$ source apportionment based on multi-model with organic tracers in Beijing during haze episodes. *Science of the Total Environment, 772*, 144766. https://doi.org/10.1016/j.scitotenv.2020.144766.

[19] Song, Y., Zhou, A., Zhang, M. (2020). Exploring the effect of subjective air pollution on happiness in China. *Environmental Science and Pollution Research, 27*(34), 43299-43311. https://doi.org/10.1007/s11356-020-10255-8.

[20] Wang, Z., Chen, H., Zhu, J., Ding, Z. (2022). Daily $PM_{2.5}$ and PM_{10} forecasting using linear and nonlinear modeling framework based on robust local mean decomposition and moving window ensemble strategy. *Applied Soft Computing, 114*, 108110. https://doi.org/10.1016/j.asoc.2021.108110.

[21] Grantz, D. A., Garner, J. H. B., Johnson, D. W. (2003). Ecological effects of particulate matter. *Environment International, 29*(2-3), 213-239. https://doi.org/10.1016/S0160-4120(02)00181-2.

[22] Naveen, V., Anu, N. (2017). Time series analysis to forecast air quality indices in Thiruvananthapuram District, Kerala, India. *International Journal of Engineering Research and Applications, 7*(6), 66-84. https://doi.org/10.9790/9622-0706036684.

[23] Mishra, D., Goyal, P. (2015). Analysis of ambient air quality using fuzzy air quality index: A case study of Delhi, India. *International Journal of Environment and Pollution, 58*(3), 149-159. https://doi.org/10.1504/IJEP.2015.077173.

[24] Kliengchuay, W., Srimanus, R., Srimanus, W., Niampradit, S., Preecha, N., Mingkhwan, R., Worakhunpiset, S., Limpanont, Y., Moonsri, K., Tantrakarnapa, K. (2021). Particulate matter (PM_{10}) prediction based on multiple linear regression: A case study in Chiang Rai Province, Thailand. *BMC Public Health, 21*(1), 1-9. https://doi.org/10.1186/s12889-021-12217-2.

[25] Aladağ, E. (2021). Forecasting of particulate matter with a Hybrid ARIMA model based on wavelet transformation and seasonal adjustment. *Urban Climate, 39*, 100930. https://doi.org/10.1016/j.uclim.2021.100930.

[26] Koutsoyiannis, D., Cohn, T. (2008). The Hurst phenomenon and climate. *EGU General Assembly 2008 Geophys. Res. Abstarcts 10.* https://doi.org/10.13140/RG.2.2.13303.01447.

[27] Chaudhuri, S., Dutta, D. (2014). Mann-Kendall trend of pollutants, temperature and humidity over an Urban Station of India with forecast verification using different ARIMA models. *Environmental Monitoring and Assessment, 186*(8), 4719-4742. https://doi.org/10.1007/s10661-014-3733-6.

[28] Abd Rahman N., Yusop Z., Şen Z., Taher S., Lawal K. I., Bahru J. (2017). Mitigation of time series approach on climate change adaptation on rainfall of Wadi Al-Aqiq, Madinah, Saudi Arabia. *Journal Teknologi, 79*(5). https://journals.utm.my/jurnalteknologi/article/view/10206.

[29] Kane, I. L., Suresh, V. M., Abubakar, H. U. (2020). Dynamical trend and homogeneity analysis of rainfall time series in Thiruvallur District, Tamil Nadu, India. *Abuja Journal of Pure and Applied Sciences, 1*(2), 7-12. https://www.ajoumr.com/index.php/ajoumr/article/view/25.

[30] Yusof, F., Kane, I. L., Yusop, Z. (2013). Structural break or long memory: an empirical survey on daily rainfall datasets across Malaysia. *Hydrology and Earth System Sciences, 17*(4), 1311-1318. https://doi.org/10.5194/hess-17-1311-2013.

[31] Granger, C. W. J., Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics, 73*(1), 61-77. https://doi.org/10.1016/0304-4076(95)01733-X.

[32] Dueker, M., Asea, P. K. (1995). Non-monotonic long memory dynamics in black-market exchange rates. *Federal Reserve Bank of St. Louis Working Paper Series (1995-003).* http://research.stlouisfed.org/wp/1995/95-003.pdf.

[33] Gil-Alana, L. A. (2009). Persistence and time trends in the temperatures in Spain. *Advances in Meteorology, 2009*, 1-8. https://doi.org/10.1155/2009/415290.

[34] Geweke, J., Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis, 4*(4), 221-238. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-

9892.1983.tb00371.x.
[35] Shimotsu, K. (2006). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory, 26*(2), 501-540. https://doi.org/10.1017/S0266466609100075.
[36] Wang, W., Van-Gelder, P. H. A. J. M., Vrijling, J. K., Chen, X. (2007). Hydrology and earth system sciences detecting long-memory: Monte Carlo simulations and application to daily streamflow processes. *Hydrol. Earth Syst. Sci., 11*(2), 851-862. www.hydrol-earth-syst-sci.net/11/851/2007.
[37] Brockwell, P. J., Davis, R. A. (2002). *Introduction to time series and forecasting*. New York, NY: Springer New York. https://link.springer.com/chapter/10.1007/0-387-21657-X_8.
[38] Box, G. E. P., Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association, 70*(349), 70-79. https://www.tandfonline.com/doi/abs/10.1080/01621459.1975.10480264.
[39] Liu, K., Chen, Y. Q., Zhang, X. (2017). An evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs. *Axioms, 6*(2), 16. https://doi.org/10.3390/axioms6020016.
[40] Caporale, G. M., Gil-Alana, L. A., Carmona-González, N. (2021). Particulate matter 10 (PM_{10}): Persistence and trends in eight European capitals. *Air Qual Atmos Health, 14*(7), 1097-1102. https://doi.org/10.1007/s11869-021-01002-0/Published.
[41] Gil-Alana, L. A., Trani, T. (2019). Time trends and persistence in the global CO_2 emissions across Europe. *Environmental and Resource Economics, 73*(1), 3-228. https://doi.org/10.1007/s10640-018-0257-5.
[42] Chai, T., Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the Literature. *Geoscientific Model Development, 7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014.
[43] Zelenka, A., Perez, R., Seals, R., Renne D. (1999). Effective accuracy of satelite-derived hourly irradiances. *Theor. Appl. Climatol. 62*, 199-207. https://link.springer.com/article/10.1007/s007040050084.