International Journal of Environmental Science and Technology
https://doi.org/10.1007/s13762-021-03672-w

ORIGINAL PAPER

Trend analysis and change point detection of air pollution index in Malaysia

Y. Alyousifi[1] · K. Ibrahim[1] · W. Z. W. Zin[1] · U. Rathnayake[2]

Received: 20 April 2021 / Revised: 6 September 2021 / Accepted: 11 September 2021

Abstract
Parametric methods are commonly used to conduct the trend analysis of air pollution. These methods require certain statistical assumptions, such as stationarity and normality of the data. However, such assumptions are usually not applicable to trends in Air pollution index (API). In addition, the change points in the time series have not been taken into consideration by most of the analysis of API. Therefore, this study presents a comprehensive investigation of the trend analysis and change point detection of the mean and maximum of API series in Malaysia. The hourly, daily, weekly, monthly, seasonal, and annual API data series were considered in the analysis. The finer time intervals were used to detect any significant increasing or decreasing trends of the API series for Malaysia. The API data were collected from 37 air monitoring stations in Peninsular Malaysia. The nonparametric tests, including Mann–Kendall test, Pettitt test, and innovative trend analysis were used to examine the contribution presented herein. Various aspects of API data were studied, including upward trends, downward trends, and change points. Several significant monotonic trends and changing points in some of the API measuring stations were found from the Mann–Kendall test results. Significant increasing trends of the monthly and seasonal mean, as well as maximum API, were found in the years 2013 and 2014 for some stations. In addition, the magnitudes of the increasing trends in maximum API are larger than the mean API. The detection points captured by the Pettitt test are possibly related to the El-Nino events. In general, the results of the study provide comprehensive information on air quality trends and their atmospheric aspects, which can help in developing strategies to address air quality problems and provide meaningful opportunities to mitigate air pollution problems in Peninsular Malaysia.

Keywords Air pollution assessment · Air pollution index (API) · Innovative trend analysis · Mann–Kendall (MK) test · Peninsular Malaysia · Pettitt test

Introduction

Air pollution is one of the main environmental challenges, which can be seen in most countries worldwide. However, air pollution is not only an environmental challenge but also an alarming environmental problem in Malaysia. Air pollution index (API) in some of the continuous air monitoring stations in Malaysia have showcased that they have reached hazardous levels in the last few years. Therefore, this is a serious concern, particularly when the weather in this part of the world is proven to experience adverse air pollution conditions. Air pollution occurrences, duration, and severity are a few of the most important variables in the field of air quality management (Gass et al. 2015; Halim et al. 2018). Therefore, the urban and industrial areas in Peninsular Malaysia are particularly the main contributors to heavy pollutant emissions, resulting in major air pollution problems throughout the region.

Air pollution in Malaysia has been investigated by many researchers (Al-Dhurafi et al. 2018a; Nieto et al. 2018; Halim et al. 2018; Motiva et al. 2019; Maleki et al. 2019; Alyousifi et al. 2018, 2019, 2020a). Similar to many other countries, the urban and industrial areas in Malaysia are considered to

Editorial responsibility: Mohamed F. Yassin.

✉ Y. Alyousifi
p90499@siswa.ukm.edu.my

1 Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

2 Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

be the most affected areas, due to dense traffic and manufacturing industries (Latif et al. 2011; Azid et al. 2014). Therefore, urbanization coupled with industrial development has contributed to the presence of high amounts of atmospheric pollutants (Azid et al. 2014; Dominick et al. 2012 Latif et al. 2011). Based on the studies by Azmi et al. (2010) and Dominick et al. (2012), crowded vehicular traffic is one of the major sources of air pollution in urban areas for most of the developing countries, including Malaysia. In addition, the open burning and forest fires, which often occur in the neighboring countries, could be another source of air pollution in Malaysia.

Particulate Matter (PM_{10}) and Ozone (O_3) are mainly contributing to the air pollution in Peninsular Malaysia. These two pollutants are the most concentrated pollutants compared to many others in the context of air pollution. For instance, PM_{10} was found to contribute as much as 70% of the total pollutants in the air, while the other 30% was due to the presence of high O_3 levels during the period of unhealthy air quality in 2002. However, the other pollutants, such as Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), and Carbon Monoxide (CO), had a slight effect on API values. Thus, PM_{10} and O_3 are the most important contributors in determining and selecting the API values. The concentration level of these pollutants reflects the healthy state of the air in the atmosphere, and therefore, actions can be taken if the air quality is poor (Afroz et al. 2003). The main sources of PM_{10} and O_3 are transboundary sources and the number of motor vehicles (Azmi et al. 2010; Mohtar et al. 2018). In addition, the main sources of SO_2 are the burning of fossil fuel at power plants and industrial facilities (Butler and Whelan 2018), and those of CO are the biomass burning and industrial activities (Zhang et al. 2009).

Research on trend analysis is quite popular among climate and environment researchers. The climatic trends are widely investigated and the results are effectively used in future planning activities in many countries. However, the trend analysis and change point detection in air pollution have been considered by only a few researchers (Wang and Zhang 2012; Zarenistanak et al. 2014; Zhang et al. 2018). Azmi et al. (2010) have conducted a study to determine the trends and status of air quality in several stations in Klang Valley, Malaysia. They also assessed the correlation of air quality with some meteorological factors at the identified stations. They have also reported that the concentrations of CO, NO_2, and SO_2 were higher in Petaling Jaya. This could be due to the presence of heavy traffic in the area. Moreover, Abd Rani et al. (2018) have presented a statistical description of the API trends in Malaysia from the years 2010 to 2015. They have identified the highest API value of 663, which had occurred in Muar on the 23rd of June 2013. This was recorded as the worst air quality record that had been experienced in Malaysia. The trend analysis of Ozone for the period 1997 to 2007 of Kemaman district has been examined by Ismail et al. (2012) using a parametric method. The results showcased that, in general, there is an increasing trend for air pollution during the dry months and a decrease in air pollution during the rainy months in Malaysia.

Malaysia is a rapidly developing country mainly due to its industrialization. The country has been affected by various episodes of temporal and regional air pollution events (Azid et al. 2014; Arman et al. 2015; Mohtar et al. 2018; Alyousifi et al. 2020a, b). Thus, it is essential to investigate the trends of API to have a better understanding of the air pollution problem in Malaysia. Most of the studies on the statistical trend analysis of air pollution in Malaysia have been conducted based on parametric statistical methods. However, as stated above, the parametric statistical methods can only be successfully used to the normally distributed independent data. Therefore, parametric methods are not the best methods to detect the trends for the time series data. Thus, the nonparametric tests are important as they do not require the normality assumptions on the time series data (Jaiswal et al. 2018). Generally, nonparametric methods have advantages in handling missing data; thus, investigators can get a more refined view of the variable tendencies in time series (Ali et al. 2019).

Nevertheless, as stated above, the nonparametric methods have not received much attention in detecting the air pollution trends in Malaysia. Thus, the need for such a study is highly appealing. Mann–Kendall and Pettitt tests are nonparametric statistical methods that are commonly used to assess the existence of trends and change points, respectively. These are widely used in the time-series data such as rainfall, temperature, and river flow (Deni et al. 2010; Damberg and AghaKouchak 2014; Yusof et al. 2013). Furthermore, to the best of authors' knowledge, the literature does not showcase the usage of graphical methods, such as innovative trend analysis methods for identification of trends in API data of Malaysia.

However, the literature shows many related studies in the nearby countries. Chaudhuri and Dutta (2014) have implemented the Mann–Kendall test to observe the trends in pollutants data, humidity, and temperature over an urban station of India. Their findings demonstrated the existence of a significant increasing trend in all the pollutant parameters and average surface temperature during the post-monsoon season. In addition, Jaiswal et al. (2018) have applied the Mann–Kendall and Sen's slope estimator tests on seasonal data for air quality index for Varanasi, India for the years 2013 to 2016. An increasing trend was found for PM_{10} while decreasing trends were found for $PM_{2.5}$, CO, NO_2, and SO_2 in Varanasi.

The innovative trend analysis technique proposed by Sen and Srivastava (2012) is a nonparametric graphical method to detect trends in time series. The test does not also depend on data distribution, autocorrelation, and length of data series (Tosunoglu and Kisi 2017). The innovative trend analysis method allows robust identification of the trends by ignoring the seasonal cycle and the length of time series. It also allows examination of any hidden sub-trends that are presented in the series. Thus, the innovative trend analysis method can be used for investigating a monotonic trend present in the time series, which persists with time. Therefore, the method has been widely applied in different studies around the world in recent years (Tabari et al. 2015; Dabanlı et al. 2016; Demir and Kisi 2016; Öztopal and Sen 2017; Ali et al. 2019). For instance, Öztopal and Sen (2017) have applied the innovative trend analysis method for assessing the rainfall trend at seven stations in Turkey. According to the literature, it is observed that innovative trend analysis had been used in different studies. However, to the best of the authors' knowledge, no study has applied innovative trend analysis on analyzing the API trend in Malaysia.

However, a number of studies on air pollution which use API data can be found in Malaysia. The API is a generalized measure of air pollution conditions for describing the air quality in the environment. API data are inherently random and thus, they do not require statistical assumptions of linearity and stationarity. Hence, API data can be effectively used to detect trends and change point detection of air pollution based on nonparametric methods. However, the trends or change points in the API data are yet to be investigated in Peninsular Malaysia.

Therefore, this study presents an answer to the above research gap by conducting a comprehensive investigation on the trend analysis with a focus on the distributional shifts and behaviors of the air pollution index in Peninsular Malaysia. The trend analysis of the API data was performed based on Mann–Kendall and the innovative trend analysis methods for identifying the air pollution trends in all air monitoring stations of Peninsular Malaysia (37 stations). The distributional shifts in API were described using the Pettitt test. In addition, the mean and maximum APIs were analyzed in order to provide details on the trend of the air pollution problem. Furthermore, the API time series analysis is discussed, and a description of the study area and descriptive statistics of API data are presented. Generally, the findings of this research are expected to assist policymakers in planning and preparing suitable strategies for managing the uncertain changes in air quality.

Study area and data collection

The hourly and daily API data for three years from 1st January 2012 to 31st December 2014 were collected for 37 air monitoring stations in Peninsular Malaysia. These stations are located in several cities and the locations of the study areas are shown in Fig. 1 (Alyousifi et al. 2020a).

The API is adopted as an indicator of air pollution conditions in Malaysia, which was designed according to the US model that numerically ranges from 0 to ∞. This index reflects the air quality levels related to human health. The API data are structured based on the information on five main air pollutant variables, namely, Ozone (O_3), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Particulate Matter (PM_{10}), and Sulfur Dioxide (SO_2) (DOE 2000). In addition, the API values are determined based on the highest index value among these five pollutants of a particular time period (up to a 24-h averaging period), as shown in Fig. 2. The main advantage of the API is that it has multiple categories that allow for temporal flexibility in the evaluation of air pollution.

A moderate state of air pollution level is denoted by an API value of less than 100, while a value of greater than 100 indicates a higher degree of air pollution. The breakpoints of API values and the respective health status, which are good, moderate, unhealthy, very unhealthy, hazardous, and emergency for Malaysia can be found in the regulations of the Department of the Environment of Malaysia (DOE 2000; Alyousifi et al. 2020b).

As stated earlier, API is categorized based on the highest values from five main air pollutant index values (PM_{10}, O_3, SO_2, NO_2, and CO), where PM_{10} and SO_2 hourly values are averaged over a 24-h running period, CO is averaged over 8 h, while O_3 and NO_2 are on one hour running averages. After that, API is calculated with the use of sub-index functions for each pollutant based on the standpoint of human health implications (DOE 2000; Alyousifi et al. 2021). The calculations of the sub-indices of pollutants can be found in the study by Abd Rani et al. (2018).

These five important parameters are measured by the sensor-based measuring devices in Malaysia. The optical particle counters detect the particulate pollution by measuring the light scattered due to particles. The optical sensors detect gases like carbon monoxide and carbon dioxide by measuring the absorption of infrared light. The mobile Air Pollutant Index (API) Monitoring System consists of a Sharp GP2Y1010AU0F optical dust detector as a sensor for dust, Arduino Uno, and LCD Keypad Shield. A signal conditioner is used to amplify and extend the range of the sensor reading for a more accurate result. These are sensitive and accurate instruments, thus, they are widely used over many air pollution stations over the world.

Materials and methods

The nonparametric methods are commonly used for analyzing the time series trends and detecting the change points because they are less sensitive to outliers and do not require

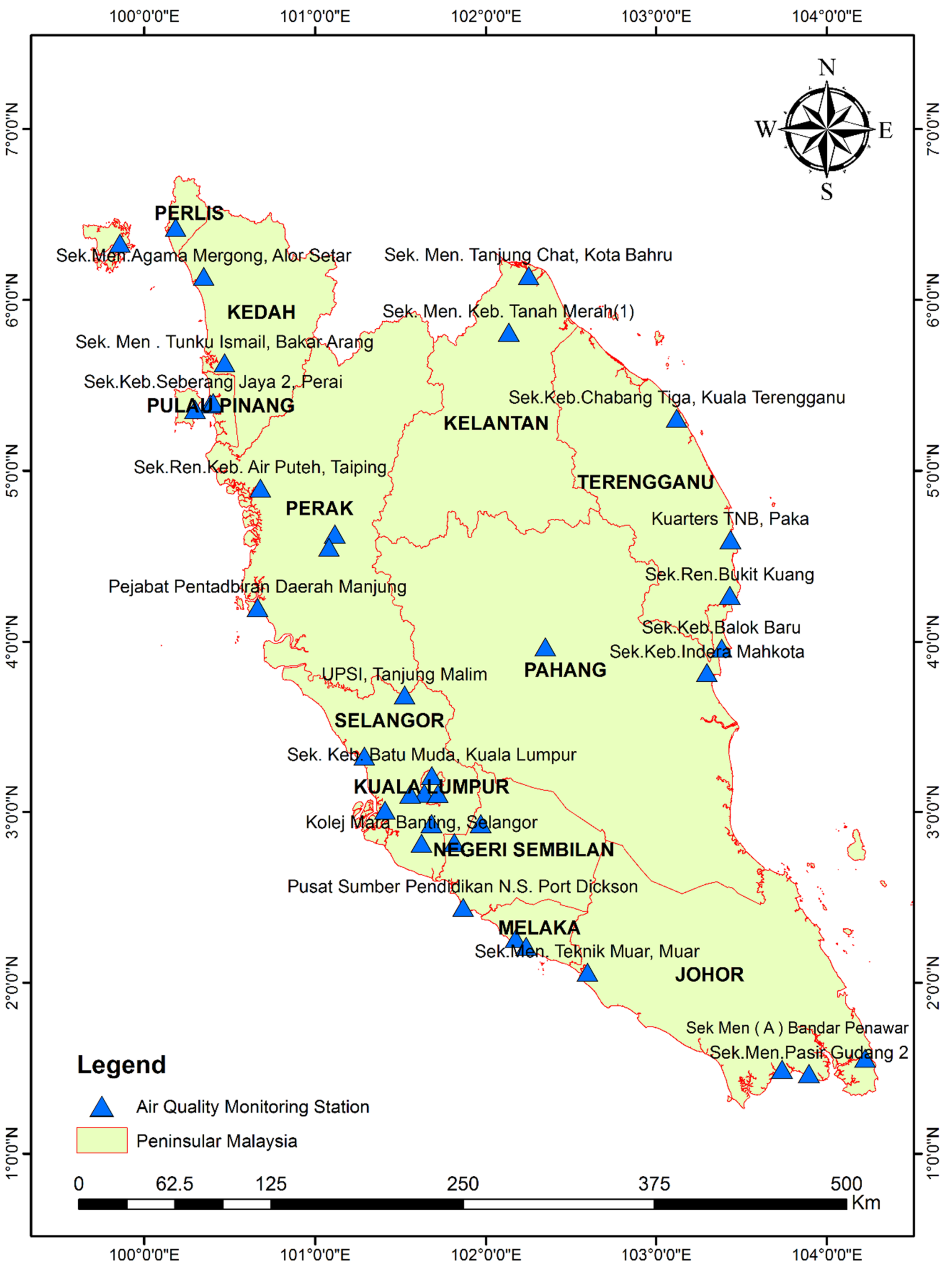

Fig. 1 Locations of 37 air quality monitoring stations in Peninsular Malaysia (Alyousifi et al. 2020a)

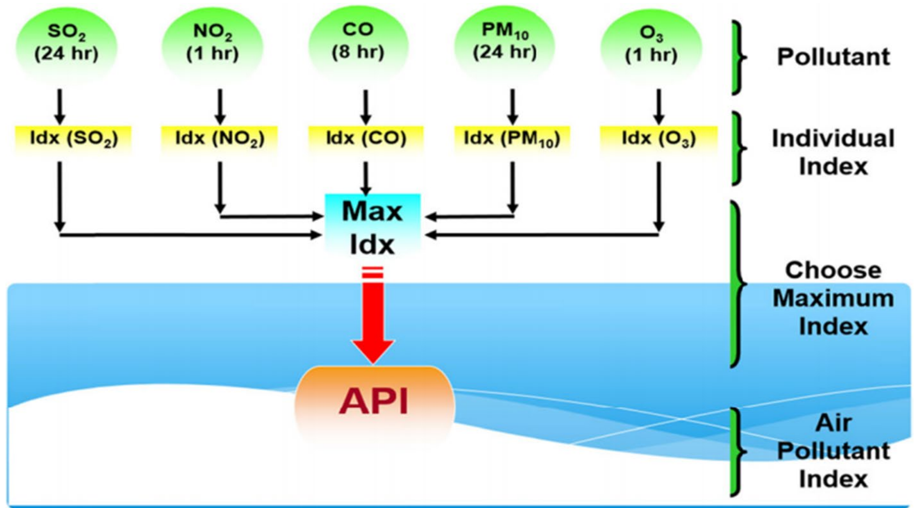

Fig. 2 A method of determination for the APIs DOE (2000)

any statistical assumption about the data, such as linearity and stationarity. Thus, this section emphasizes the theories behind the overall trend and change-points of the data series at each station. The nonparametric methods such as the Mann–Kendall test, Sen's slope, and innovative trend analysis methods are usually used for trend analysis, while the Pettitt test is widely used for change point detection. These tests are well recommended by the World Meteorological Organization for public application (Mitchell et al. 1994; Jaiswal et al. 2015; Suhaila and Yusop 2018). The purpose of this study is to investigate the presence of possible trends and investigate the variability of API at different time points in order to present a better understanding of the behavior of air pollution in Peninsular Malaysia over time. In addition, examining the variability of API for a better understanding of the stochastic behavior of air pollution is carried out.

The Mann–Kendall (MK) test

The Mann–Kendall (MK) test (Mann 1945 and Kendall 1975) is applicable for detecting a monotonic trend in a time series with outliers. The statistics of the tests are based on the sign of differences, but not directly based on the values of the variable (Pal and Al-Tabbaa 2009). Identifying the significant trends and determining changes in a trend over time are very vital for the analysis of API time series. The MK test can successfully be used to statistically assess the presence of a monotonic trend, either increase or decrease, in a time series (Suhaila and Yusop 2018).

The test statistic of the Mann–Kendall trend test is calculated according to the following equations (Eqs. 1–5).

$$S_{mk} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}\left(x_j - x_i\right) \tag{1}$$

where n is the length of data, x_i and x_j is the two generic sequential data, and $sgn\left(x_j - x_i\right)$ can be written as

$$\mathrm{sgn}(x) = \begin{cases} 1 : x_j - x_i > 0 \\ 0 : x_j - x_i = 0 \\ -1 : x_j - x_i < 0 \end{cases} \tag{2}$$

The test statistic S_{mk} is asymptotically normally distributed, and under the null hypothesis it has an expectation $E(S_{mk})=0$ and variance $V\left(S_{mk}\right)$ is given by

$$V\left(S_{mk}\right) = n(n-1)(2n+5)/18 \tag{3}$$

in the case of that there is no tie in the data (Mann 1945; Kendall 1948). Otherwise, it can be given by

$$V\left(S_{mk}\right) = \frac{1}{n}\left[n(n-1)(2n+5) - \sum_{t} t_j\left(t_j - 1\right)\left(2t_j + 5\right)\right] \tag{4}$$

where t is the number of data points in the jth tied group. For this test, the statistics Z can be computed as

$$Z = \begin{cases} \frac{S+1}{\sqrt{V(S_{mk})}} & : S > 0 \\ 0 & : S = 0 \\ -1 & : S < 0 \end{cases} \tag{5}$$

Having no ties in data, the well-known Kendall's τ is also calculated as S_{mk}/n^2, which is a measure of rank correlation (Pohlert et al. 2016; Militino et al. 2020). The null hypothesis is usually tested at 5% significance level. A positive value of Z indicates an increasing trend and a negative value of Z indicates a decreasing trend. In order to apply the Mann–Kendall test, a whitening test is performed on data to check for the possible correlation in the series. The aim is to identify the presence of any autocorrelation effect in the time series data, which may influence the Mann–Kendall test's ability to evaluate the significance of the trend (Von Storch 1999).

Sen's Slope estimator

The magnitude of the trend, which is the slope of the monotonic trend, can be estimated using Sen's slope estimator test. The simple nonparametric procedure is given by Eq. 6.

$$d_i = \text{median}\left[\frac{y_j - y_i}{j - i}\right], \quad \text{for all } j > i \tag{6}$$

where y_j and y_i are the sequential data values at times j and i, respectively (Theil 1950; Sen 1968; Hirsch et al. 1982). Also, $1 < K < j < n$, and d_i are considered a median of all possible combinations of pairs for the whole data set (Salarijazi et al. 2012). The alternative hypothesis is that a trend exists, and this trend can be either positive or negative.

Graphical method in trend analysis

Usually, some API series are not serially independent. Thus, API data show a significant serial correlation. The errors can be occurred in detecting trends even for a moderate correlation in the API series (von Storch 1999). Even though the procedure of pre-whitening is applied to address the issue of serial correlation to the API series, the null hypothesis is also rejected at some API series. Therefore, an innovative trend analysis (ITA) method, also known as a graphical method in trend analysis, was employed to identify the positive and negative trends in API data. Although this method is widely used in identifying trends of climate series (Rathnayake 2019), it has not been applied for API series yet. The following steps are carried out for implementing the innovative graphical method.

1. First, divide the data series into two equal subsets
2. Then, arrange the sub-data sets in ascending order.
3. Next, plot two antecedent series in the Cartesian coordinate system (the older one in the *X*-axis and the recent one in the *Y*-axis) and draw a 45° line and two other lines at + 5% and − 5% at the same coordinate system.

If the time series is between + 5% and − 5%, the test presents a "no trend" situation. However, if the data scatter is placed above + 5%, a positive trend in the recent data compared to the older data can be predicted and vice versa. The ± 5% lines can even be extended to ± 10% lines for no trends and the scatters can also be grouped in identifying the partial trends (Şen 2017; Öztopal and Şen 2017).

Pettitt test

The Pettitt test (Pettitt 1979) is applied in order to test for a shift in the central tendency of a time series. Suppose that the random variable, $X_i, i = 1, 2, \dots, n$ is independent and identically distributed and assume that the X_i's have the same mean in the data series and the series is homogeneous. The "no change" is the null hypothesis (H_0) and the "change points exist" is the alternative hypothesis (H_a). If *P* value is less than 0.05, then, the change point is significant (Pohlert et al. 2016; Suhaila and Yusop 2018).

The governing equations for Pettitt's test are shown in Eqs. 7 and 8.

$$S_P = \max_k \left| \sum_{i=1}^{k} \sum_{j=i+1}^{n} \text{sgn}\left(x_j - x_i\right) \right|, 1 < k < n \tag{7}$$

The change-point of the series is located at S_P, provided that the statistic is significant. The significance probability of S_P is approximated for $p \leq 0.05$ with;

$$P_p = 2\exp\left[\frac{-6S_P^2}{n^2 + n^3}\right] \tag{8}$$

The distribution of S_P is symmetric around zero with $E(S_P) = 0$ under the null hypothesis and for each k. It is expected to have large values for S_P, when there is a shift in data (Pettitt 1979; Pohlert et al. 2016; Militino et al. 2020).

Overall methodology

The obtained hourly and daily API data sets for 37 air quality monitoring stations were tested using the Mann–Kendall test and Sen's slope tests. However, the tests were carried out in daily, monthly, and seasonal time series. The two main seasons in which the air quality can be affected in Malaysia are the Southwest Monsoon (SWM) and the Northeast Monsoon (NEM) seasons. These monsoons occur in the months of May to August and November to February, respectively. The country also experiences two inter-monsoon seasons periods from March to April and September to October, respectively. The amount of rain is generally larger in the monsoon periods as compared to the inter-monsoon periods. Therefore, NEM and SWM were mainly considered for the time series analysis of API. In addition, the trends in APIs were tested based on the innovative trend analysis technique. Furthermore, the change points of the API data series for each station were investigated using the Pettitt test. The results of the analyses are presented and discussed in the next section.

Results and discussion

API data representation

The data used in this study consist of approximately 26,304 hourly observations at each station for a total of 37 stations. The descriptive statistics and time series analysis of API data have been introduced before the investigation of the trend analysis and change point detection of API. Table 1 provides a summary of the descriptive statistics for the daily API data for each station.

The mean values for all cities are greater than their corresponding medians, which indicates that the API distributions are positively skewed or right-skewed. This observation is clearly shown in Fig. 3. Figure 3 displays histogram plots of API data for the most polluted cities in Peninsular

Table 1 Some descriptive statistics on the API data of air quality monitoring stations in this study

Station ID	Location of Stations	Mean	Max	Min	SD	Skewness	Kurtosis
CA0001	Johor	54.41	**352**	20	19.78	7.21	**96.32**
CA0002	Terengganu	55.85	262	22	14.91	3.89	41.76
CA0003	Pulau Pinang	50.70	120	18	16.00	0.65	0.37
CA0006	Melaka	63.54	**477**	17	**24.05**	9.36	133
CA0007	Pahang	39.56	138	15	12.71	1.33	5.68
CA0008	Perak	56.74	109	29	13.88	0.55	0.08
CA0009	Pulau Pinang	53.82	132	20	14.61	1.27	2.83
CA0010	Negeri Sembilan	60.82	177	26	16.69	2.14	9.67
CA0011	Selangor	65.16	**495**	25	**30.29**	7.41	85.1
CA0014	Pahang	46.21	172	16	12.89	1.97	13.8
CA0015	Pahang	49.63	200	2	13.72	2.61	24.2
CA0016	Selangor	57.39	231	26	18.53	2.94	18.8
CA0017	Kedah	56.03	121	25	14.66	1.16	2.27
CA0019	Johor	51.17	226	17	18.04	3.14	20.3
CA0020	Perak	46.48	155	16	16.36	1.45	4.54
CA0022	Kelantan	44.25	116	11	11.88	0.69	2.43
CA0024	Terengganu	38.05	162	5	13.43	1.93	11.3
CA0025	Selangor	63.21	**301**	7	21.96	2.39	15.9
CA0032	Kedah	40.36	102	15	10.63	1.22	3.24
CA0033	Perlis	43.04	104	15	12.37	0.53	0.34
CA0034	Terengganu	49.77	140	0	12.51	0.67	4.54
CA0038	Pulau Pinang	56.63	121	18	16.30	0.53	0.24
CA0040	Kedah	43.80	108	17	13.77	1.11	2.18
CA0041	Perak	44.80	**333**	14	19.66	4.31	47.6
CA0043	Melaka	54.52	**415**	16	**23.40**	6.50	78.1
CA0044	Johor	53.02	**663**	7	**28.10**	12.9	**247**
CA0045	Perak	53.09	179	11	21.95	1.41	2.59
CA0046	Perak	62.00	247	27	15.32	2.76	24.7
CA0047	Negeri Sembilan	58.26	173	23	19.42	1.17	3.31
CA0048	Selangor	45.32	247	14	21.10	2.88	16.1
CA0053	W.P. Putrajaya	56.13	223.5	19	19.80	2.06	10.1
CA0054	Kuala Lumpur	68.93	186	26	23.01	1.04	1.90
CA0056	Negeri Sembilan	60.41	**364**	26	**22.06**	5.34	58.2
CA0057	Johor	51.75	**327**	25	19.98	5.98	68.3
CA0058	Kuala Lumpur	66.95	200	21	23.82	1.05	2.45
CA0059	Kelantan	51.01	134	7	15.62	0.24	2.51
CA0060	Selangor	66.65	**323**	30	23.92	4.13	34.2

The highest values of API are represented in bold

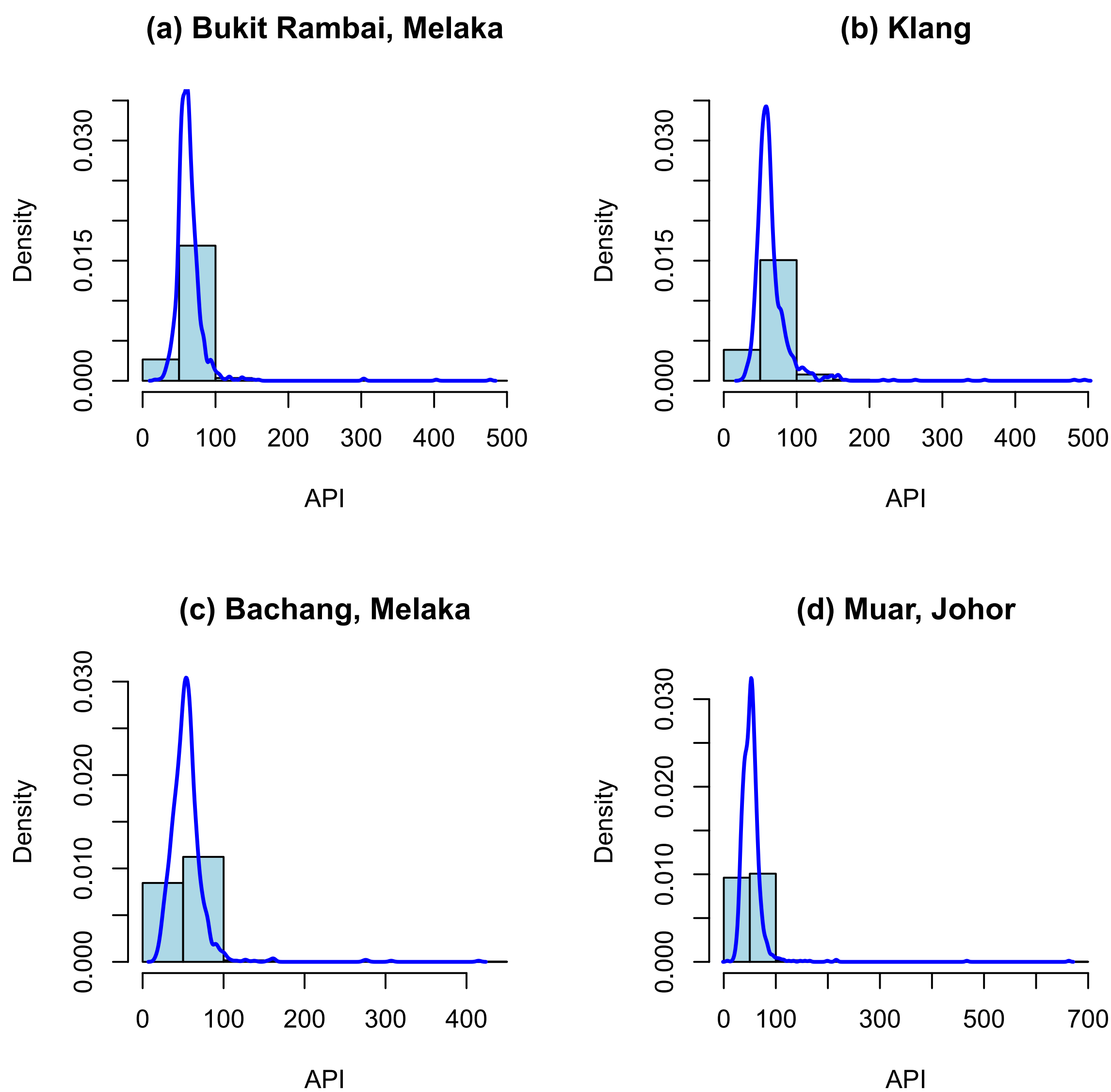

Fig. 3 Histogram of API data for the most polluted stations in Peninsular Malaysia

Malaysia (Klang, Bukit Rambai in Melaka, Bachang in Melaka, and Muar in Johor).

The majority of API data are concentrated around values less than 100 for these stations. However, it can also be seen that a small number of API values is placed beyond 100. These extreme API values are contributing to moving the data distribution to the right, known as right skewed.

Some of the locations have recorded hazardous state values (API > 500) and they can be clearly seen in Table 1. The highest API observed was 663 and at a value of skewness of 247. This was recorded at Muar in Johor from 01:00 h to 16:00 h on the 23rd of June 2013. This demonstrates that Muar had experienced higher API compared to the other cities or towns during the study period. Peninsular Malaysia was facing a serious air quality problem due to the occurrence of haze during the months of June 2013 and March 2014. Several other stations can also be observed with very high API values apart from the station of Johor. They are Klang (API of 495), Bachang Melaka (API of 477), Bukit Rambai Melaka (API of 415), and Port Dickson Negeri Sembilan (API of 364) stations. These API readings also approach the hazardous state of air pollution. Therefore, these five stations can be considered the most polluted stations in Peninsular Malaysia during this period, where three of these five stations are classified as urban stations, while the other stations are industrial and sub-urban. This implies that most of the air-polluted stations in Peninsular Malaysia are urban stations. Therefore, it can be clearly understood that urbanization is one of the main factors for the adverse air quality in Peninsular

Malaysia. This is justifiable when the urban areas are surrounded by various industries with burning fossil fuel, power stations, municipal and industrial waste dumps, and incinerators. In addition, the heavy vehicular motions would worsen the air quality in the urban areas compared to the rural areas.

Additionally, the high API values found in the most polluted stations in the Peninsula, (as shown in Fig. 3) are believed to be caused by episodes of haze, due to transboundary pollution from forest fires in Indonesia (Rahman et al. 2015; Abd Rani et al. 2018). Thus, Johor, Klang, Bachang Melaka, Bukit Rambai Melaka, and Port Dickson Negeri Sembilan stations can be considered the proper choices of the reference stations for comparison, when the interest is on highly polluted areas. The standard deviations of API for these five stations are 28.10, 30.29, 24.05, 23.40, and 22.06, respectively. Therefore, they indicate a higher variability in the API as compared to other stations. These greater variations at the stations can be attributed to the bigger differences between the maximum and minimum values of API.

Table 2 showcases the seasonal mean API observed during the SWM and NEM seasons, and the monthly mean API with the coefficient of variation (CV) for some of the selected stations (*other stations are not shown here due to the page limitations*). The CV is a statistical measure of the relative dispersion of the data points in a time series around the mean. It is obvious that the high mean API is dominant during the SWM season recorded at Kuala Lumpur (CA0058) station. The lowest mean API was recorded at Jerantut, Pahang (CA0007) station during the NEM season. In terms of variation in the API series, most stations have recorded large variability in each season, with the largest variability of approximately 30% at Muar, Johor (CA0044) station and nearly 35% at Kuala Selangor (CA0048) station during SWM and NEM seasons, respectively. Therefore, NEM season has more variation relative to its mean in the API series. On the other hand, the rest of the considered stations recorded lower variability in the seasonal mean series.

The API data have skewed distributions to the right with a long tail for high concentrations (Singh et al. 2001; Al-Dhurafi et al. 2020). Thus, the normal distribution is not a suitable way to describe air pollution data. Taking logarithms of the data is one way to overcome this difficulty, thus the data are transformed to a normal distribution. Therefore, the Paired samples Wilcoxon test (also known as Wilcoxon signed-rank test (WSRT)) can be employed to assess whether two dependent samples were selected from populations having the same distribution. The Paired samples Wilcoxon test (WSRT) is a nonparametric statistical test that is utilized to compare two samples. The purpose of this test is to investigate the changes of ranks of population means. It can be used as an alternative to the paired Student's t test, when the difference between the two-sample means cannot be assumed to be normally distributed.

The difference between the two seasons was statistically significant for most of the stations according to WSRT. This can be clearly understood from the *P* values listed in Table 2, which are below the 5% significance level. This implies that the air pollution events that occurred during the SWM season are different from those occurred during the NEM season. This could be due to the extent of dryness that was resulted from the El-Nino phenomenon. This can possibly be contributed to changes in the rainy season to the dry season, thus further escalating the problem of air pollution.

Furthermore, Supplementary Table S1 displays the maximum monthly and maximum seasonal API in the SWM and NEM seasons with their coefficient of variations. It is observed that almost all the maximum values of API observed at the station were recorded during the SWM season. In addition, it is interesting to note that in general, the stations located in the west and northwest regions of the country have higher seasonal maximum APIs compared to other regions of the country.

Northeast monsoon brings heavy rainfall and winds to the Titiwangsa Range in the western and northwestern areas of the peninsula. Thus, these areas in the west of Malaysia may be less polluted than those located in the eastern part, particularly during the NEM season (Sohaila and Yusop 2018). Nevertheless, more air pollution events occurred during the SWM season than the NEM season. This implies the existence of an indirect association with the dry season, which mostly comes from the occurrences of the El-Nino

Table 2 Statistics of seasonal **mean** API series for some of the selected studied stations

Station ID	Southwest Monsoon (SWM)		Paired WSRT *P* value	Northeast Monsoon (NEM)		Monthly API	
	Mean	CV		Mean	CV	Mean	CV
CA0007	46.94	0.095	**0.0001**	**32.29**	0.162	39.69	0.214
CA0044	58.08	**0.292**	0.0637	47.95	0.210	53.28	0.259
CA0048	51.06	0.249	**0.0479**	41.15	**0.350**	45.61	**0.292**
CA0058	**72.61**	0.098	**0.0015**	60.36	0.195	67.26	0.164

Significant values are represented in bold

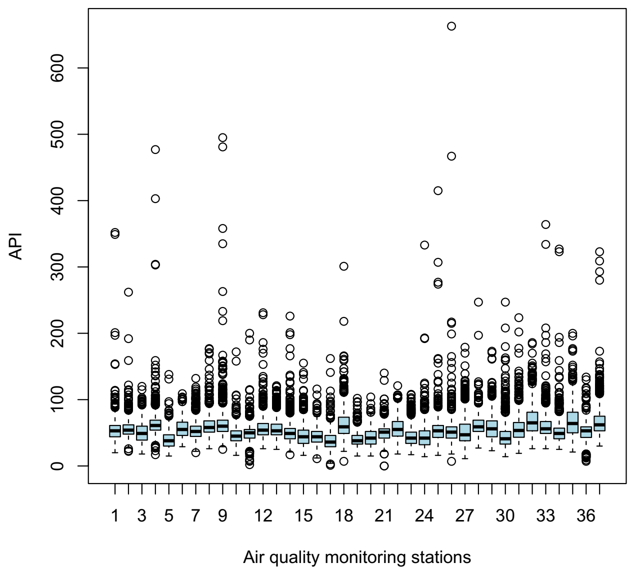

Fig. 4 The trend of daily API (for 2012–2014)

phenomenon. In addition, almost all stations were found to have the largest variability during both seasons. The largest variability of approximately 124% was found at Muar, Johor station (CA0044), while about 65% of the variability was found at Klang station (CA0011) during SWM and NEM season, respectively.

Analysis of API data

The daily API data were analyzed using statistical methods based on the R statistical software. The analysis was carried out on 1096 × 37 observations from all air quality monitoring stations in Peninsular Malaysia. Figure 4 shows the boxplots of API values from 2012 to 2014 for the 37 air quality stations. As stated above, the API values exceeding 100 can either be classified as unhealthy, very unhealthy or in hazardous status. The trend of the daily API data indicates that the majority of air monitoring stations in Peninsular Malaysia are at a good and moderate state of air pollution in general. However, some stations reached a hazardous state. Nevertheless, as already stated, the Peninsula was facing a serious air quality problem due to the occurrence of haze during this period. In particular, it is observed in Fig. 4 that nine stations have experienced high API values, which have reached hazardous status. However, twenty-eight stations have experienced lower API values, indicating that the air quality at most stations has not reached the hazardous and very unhealthy status of air pollution. These 28 stations fall within the healthy limits of humans. Also, the boxplot represents the median, which provides a helpful measure of the center of a dataset. By comparing the median to the mean, for each station, an idea of the distribution of a dataset can be obtained. It is observed that the medians are differed from the mean. Thus, a skewed distribution of the data is indicated. Therefore, this observation supports the results shown in Fig. 3.

Figure 5 shows the time series plots of the API observed at the four most polluted stations in the Peninsula (Klang; Bukit Rambai, Melaka; Bachang, Melaka and Muar, Johor). The figures clearly showcase some occurrences of unhealthy air pollution events. Therefore, the air quality in these areas seems to be a matter of concern and should be taken into account.

In addition, it could be seen from Fig. 5 that the stations have experienced the highest API levels in June 2013. The highest API was observed at the station of Muar, Johor on 23rd June 2013 (API of 663). The southwest monsoon is in its activation during the month of June. However, it was found that the southwest monsoon was a dry season due to the occurrence of El-Nino. Therefore, this had further aggravated the unhealthy air quality in the area. Also, some stations have experienced a high level of API in March 2014. Careful observation showcased that less rainfall was received at these stations as compared to other months (Alyousifi et al. 2020a, b).

Figure 6 shows the time series plots of the observed APIs (> 100) for all stations in Peninsular Malaysia. A clear trend or patterns of seasonality cannot be seen from Figs. 5 and 6. However, the non-stationarity can be seen in the time series plots for the observed API values, since these data fluctuate without the constant mean and variance. Figure 5 further indicates serious air quality issues from days 536 (19/06/2013) to 543 (27/06/2012). Severe episodes of haze have been experienced in this period. In addition, most of those stations are urban stations.

Furthermore, Fig. 6 shows the occurrences of unhealthy air pollution events (API > 100). In general, this can be understood as a warning to the whole of Malaysia on its air quality levels. Nevertheless, only a few cases can be found with higher API values greater than the MAAQG limit of 150 μg/m^3.

Figure 7 shows the histogram of APIs for all the stations considered in terms of (a) the empirical densities and (b) frequencies of API values, respectively. It can be observed that distributions of API are positively skewed. This indicates that the majority of the API data are concentrated on the left of the figures with a small number of extreme API values, as depicted by the tail of the distribution.

Trend analysis and change-point detection of air pollution index

Trend analysis of API

The existence of serial correlation in the time series was checked prior to the trend analysis of the API time series

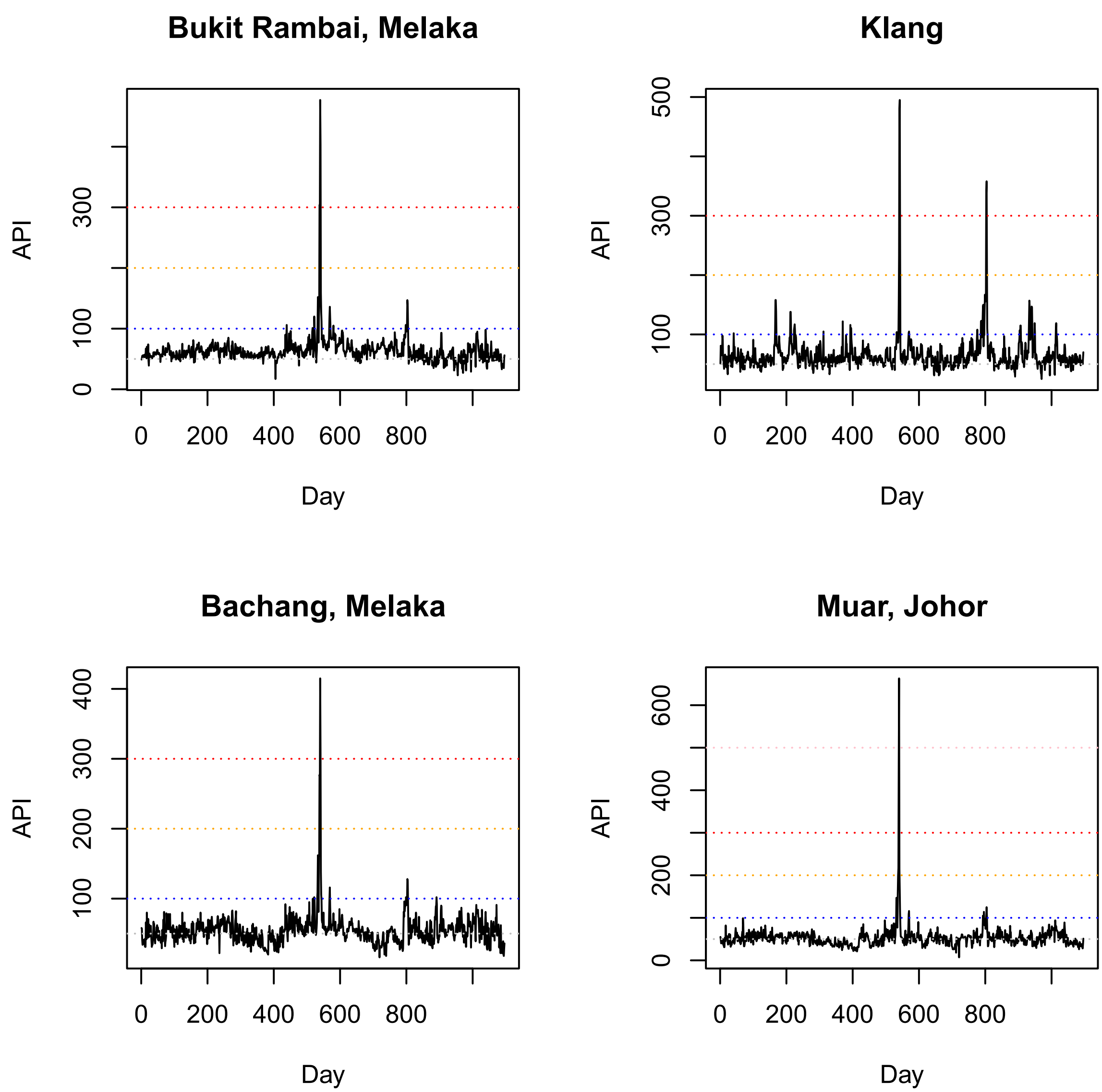

Fig. 5 Time series plots for the API values for the most polluted cities in Peninsular Malaysia

since it can affect the significance of trends. The presence of a positive serial correlation could increase the possibility of rejecting a null hypothesis of no significant trends (Douglas et al. 2000; Suhaila and Yusop 2018). Thus, it is important to check for the existence of serial correlation in the time series. The existence of autocorrelations in time series data can be determined using the one-way correlation coefficient test.

The von Neumann ratio (RVN) test (von Storch 1999) can be used to detect any presence of lag-one serial correlation using the rank. However, the presence of serial correlation among the daily API concentration levels can be visually investigated by the plot of the autocorrelation function, which computes the value of autocorrelation for the time series of daily API concentrations. The API data were tested at the 5% significance level for studying the presence of serial correlation. Most of the API series were found to be serially correlated. Figure 8 shows the plots of trend and autocorrelation of the API time series for each year, which reveals the presence of trends in the data. More results are given in Supplementary Table S2. The Von Neumann ratio (RVN) test was carried out using the "EnvStats" Package in R software (Millard et al. 2020).

It can be seen that the monthly mean API series in most of the stations are serially correlated (P value $<$ 5%). However, some stations were found to be serially independent. Nevertheless, the seasonal mean API series at all stations are serially correlated, except for four stations in SWM season and five stations in NEM season. The API series for the inter-monsoon season has not been analyzed directly.

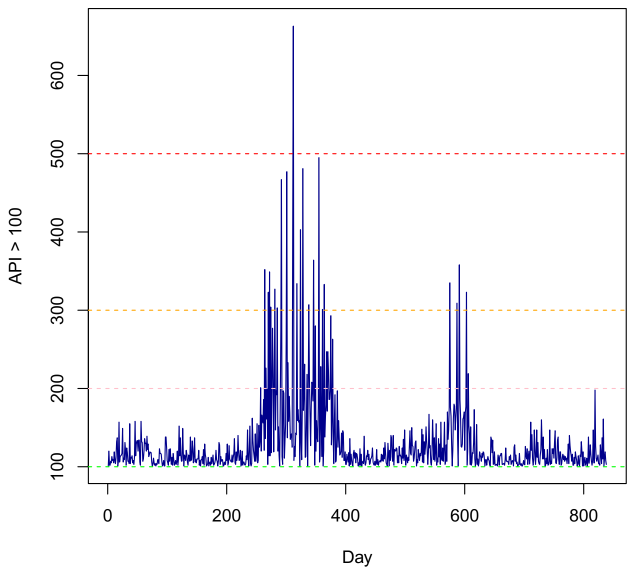

Fig. 6 Time series plot of APIs (> 100) from 2012 to 2014 in Peninsular Malaysia

However, the analysis was done indirectly using the monthly means and maximum APIs. Therefore, some results can be implemented for inter-monsoon seasons (Table 3).

In addition, Supplementary Table S2 shows that about half of the stations are serially correlated for the maximum API series. However, in contrast, serial correlation cannot be found for the seasonal maximum API in all considered stations. Nevertheless, the test suggested that the autocorrelation in these five series is not significant. However, pre-whitening was applied before the trend analysis using the Mann–Kendall test. This procedure involves estimating the slope of the monotonic trend for any series which are serially correlated. These calculations have been carried out using the "modifiedmk" Package in R (Patakamuri et al. 2020).

As previously stated, the Mann–Kendall (MK) test was used to produce trend statistics of the API series, while the magnitude of the trends in API was found by the Sen's slope estimator test. All the results of trend analysis and change points have been done based on the "Trend" and "modifiedmk" packages in R software (Pohlert et al. 2016; Patakamuri et al. 2020). The mean and maximum API series were used in the analysis. The results of trend analysis for the monthly and seasonal mean of API for all monitoring stations are shown in Table 4. The statistical significance of 5% was considered in the test.

The computed statistics of the MK test and the associated *P* values for some of the selected stations are given in Table 4. The trends in MK test were identified using the *P* value of 5%. Most of the air monitoring stations were not found with significantly increasing trends in the mean monthly API series. Nevertheless, significant positive trends in the mean monthly API series were shown by only a few stations. Few stations were found with the trends in seasonal mean API series. They are CA0024 (Paka) and CA0059 (Tanah Merah) stations for the SWM season and CA0007 (Jerantut), CA0017 (Arang), CA0022 (Kota Bahru), CA0034 (Kuala Terengganu), CA0059 (Tanah Merah), and CA0060 (Kolej Mara Banting) for the NEM season. Sen's slope values in Table 5 capture the magnitude of API trends.

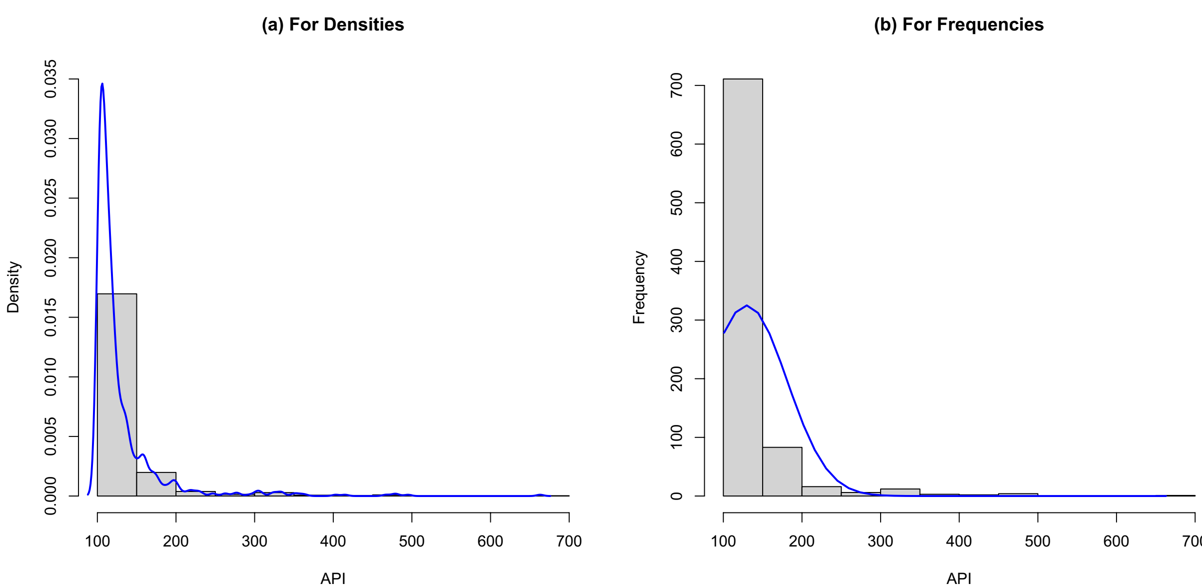

Fig. 7 Histogram of API concentration

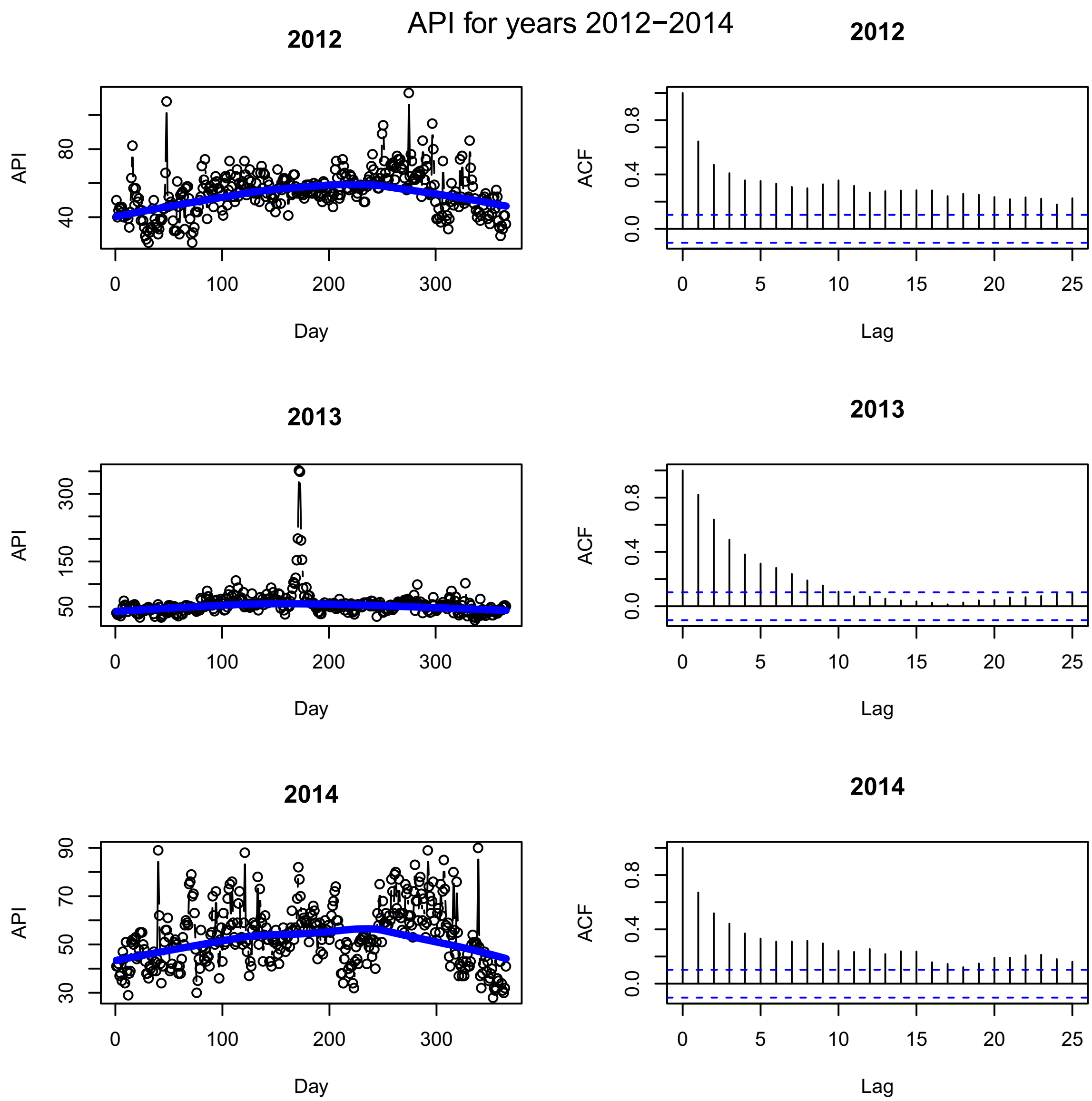

Fig. 8 Trend and autocorrelation plots of daily API for each year

Generally, increasing trends are higher than decreasing trends.

Furthermore, significant increasing trends exist in the maximum monthly API series for only five stations, which are Prai (CA0003), Paka (CA0024), University Sains Malaysia (CA0038), Kuala Lumpur (CA0058), and Kolej Mara Banting (CA0060) stations. The increasing trends of the other stations are not significant as shown in Supplementary Table S3. Also, non-significant decreasing trends are observed in the maximum monthly API series only for five stations, which are Bukit Kuang (CA0002), Klang (CA0011), Indera Mahkota (CA0014), and Cheras (CA0054) and Tanah Merah (CA0059) stations. In addition, no significant increasing trends exist in the maximum seasonal API series during SWM seasons. However, significant trends are observed for Bukit Rambai (CA0006), Paka (CA0024), Shah Alam (CA0025), and Kolej Mara Banting (CA0060) stations for the maximum seasonal API series e during the NEM seasons. Additionally, it can be clearly seen that the decreasing trends detected in the maximum monthly API series are few compared to the mean monthly series (decreasing trends are found in only five stations). However, 13 stations (34% of total stations) give decreasing trends for the maximum seasonal API series. In addition, more significant increasing trends can be found in the maximum NEM season than those in the SWM season. Therefore, the increasing trends in the monthly maximum API can be attributed to several increasing trends in the monthly and seasonal API. The Sen's slope

Table 3 Serial correlation effect on monthly and seasonal mean of API series from all stations in Peninsular Malaysia

Station ID	Monthly mean API			Mean SWM			Mean NEM		
	ρ	RVN	*P* value	ρ	RVN	*P* value	ρ	RVN	*P* value
CA0001	0.321	0.9274	**0.0005**	0.051	1.2647	0.1256	0.11	1.7179	0.5808
CA0002	0.440	0.8801	**0.0002**	−0.30	2.2559	0.6057	0.20	1.1643	0.0879
CA0003	0.289	1.3212	0.0358	−0.02	1.7941	0.6783	0.254	1.450	0.2747
CA0006	0.570	0.8839	**0.0002**	0.570	0.800	**0.0072**	0.269	1.2714	0.1416
CA0007	0.69	0.4776	**0.0000**	−0.08	1.8176	0.7135	0.344	1.0286	**0.0432**
CA0008	0.31	1.4730	0.1075	−0.04	2.0529	0.9153	0.436	1.1491	0.0817
CA0009	0.09	1.6878	0.3459	−0.12	1.9676	0.9482	0.267	1.3214	0.1731
CA0010	0.26	1.3833	0.0578	−0.07	1.9029	0.8453	0.344	1.3857	0.2202
CA0011	0.12	1.5313	0.1538	−0.07	2.1029	0.8360	0.091	1.3964	0.2287
CA0014	0.41	1.0111	**0.0014**	0.003	1.7471	0.6098	−0.08	2.0571	0.9114
CA0015	0.40	1.2726	**0.0238**	−0.37	2.7382	0.1240	0.424	0.8678	0.0154
CA0016	−0.02	1.9967	0.9920	−0.15	2.1618	0.7448	0.141	1.8357	0.7486
CA0017	0.47	0.9243	**0.0004**	0.106	1.4824	0.2899	0.628	0.6428	**0.0022**
CA0019	0.43	0.8385	**0.0001**	0.117	1.2882	0.1391	0.131	1.492	0.3144
CA0020	0.44	1.0425	**0.0022**	0.290	1.4206	0.2341	0.463	0.9892	**0.0343**
CA0022	0.17	1.6983	0.3627	−0.14	2.0794	0.8732	0.185	2.0679	0.8948
CA0024	0.47	0.9086	0.0003	0.118	1.050	**0.0421**	0.229	1.3071	0.1637
CA0025	0.17	1.4929	0.1219	−0.13	2.0618	0.9012	0.201	1.5821	0.4106
CA0032	0.29	1.2430	**0.0183**	0.098	1.3618	0.1878	0.117	1.4143	0.2435
CA0033	0.46	1.1014	**0.0043**	−0.34	2.4353	0.3760	0.413	1.0071	0.0382
CA0034	0.62	0.7060	**0.0000**	0.168	1.8022	0.6903	0.564	0.5964	**0.0013**
CA0038	0.33	1.3866	0.0593	−0.29	2.2706	0.5849	0.182	1.5036	0.3260
CA0040	0.45	1.088	**0.0037**	0.232	1.4853	0.2928	0.321	1.250	0.1294
CA0041	0.36	1.1034	**0.0044**	2.150	0.034	0.7628	0.311	1.5821	0.4106
CA0043	0.51	0.8597	**0.0002**	0.334	0.7647	**0.0053**	0.283	1.3929	0.2258
CA0044	0.27	1.4288	0.0800	0.166	1.1676	0.0796	0.268	1.4714	0.2946
CA0045	0.50	0.9405	**0.0006**	0.101	1.8059	0.6958	0.287	1.775	0.6603
CA0046	0.29	1.4813	0.1134	0.075	1.8265	0.7268	0.174	1.2464	0.1274
CA0047	0.51	0.8898	**0.0003**	0.419	1.2118	0.0986	0.295	1.3464	0.1905
CA0048	0.27	1.1822	**0.0102**	0.065	1.8559	0.7719	0.163	1.2964	0.1568
CA0053	0.34	0.8988	**0.0003**	−0.08	1.8294	0.7313	0.211	1.5929	0.4230
CA0054	0.37	1.2332	**0.0167**	0.460	0.9352	**0.0203**	0.100	1.7286	0.5954
CA0056	0.36	1.3135	**0.0336**	0.307	1.7588	0.6267	0.208	1.5571	0.3825
CA0057	0.37	1.0749	**0.0032**	0.226	1.2441	0.1146	−0.24	2.1536	0.7646
CA0058	0.45	0.8767	**0.0002**	−0.23	2.4029	0.4133	0.312	1.4464	0.2715
CA0059	0.60	1.0340	**0.0019**	0.403	1.3265	0.1632	0.467	1.0357	**0.0450**
CA0060	0.37	1.5604	0.1818	0.084	2.3324	0.5013	0.273	1.3821	0.2173

Significant values are represented in bold

results are given in Supplementary Table S3 to showcase the magnitude of the identified trends.

Mann–Kendall statistics and Sen's slope results for daily API concentrations for each year of the study period are shown in Supplementary Table S4. It is observed that significant increasing trends are presented in the daily API series for each year. It can also be observed that the significant increasing trends in the API series for most stations are fewer than the significant decreasing trends. However, the number of significant increasing trends in the daily series is greater than those found in the monthly and seasonal series. Therefore, the increasing trends in the daily mean API can be attributed to a large number of increasing trends in the monthly and seasonal API.

Table 4 Mann–Kendall statistics and Sen's slope for a **monthly** and seasonal **mean** of the API concentrations

Station ID	Monthly API mean			Southwest monsoon (SWM)			Northeast monsoon (NEM)		
	Z	Sen's slope	Sig/In-sig	Z	Sen's slope	*P* value	Z	Sen's slope	Sig/In-sig
CA0007	1.44	0.23	In-sig	−0.3	−0.16	0.77	1.98	0.69	**Sig**
CA0017	3.61	0.44	**Sig**	0.99	0.74	0.32	3.27	1.29	**Sig**
CA0022	2.08	0.18	**Sig**	−0.2	−0.14	0.84	1.937	0.74	**Sig**
CA0024	2.90	0.358	**Sig**	2.47	0.68	**0.01**	1.19	0.39	In-sig
CA0034	−2.06	−0.26	**Sig**	−0.4	−0.08	0.69	−2.82	−1.18	**Sig**
CA0059	−3.39	−0.50	**Sig**	−1.88	−1.25	**0.050**	-2.18	-0.84	**Sig**
CA0060	2.57	0.33	**Sig**	0.4	0.16	0.69	2.47	0.99	**Sig**

Significant values are presented in bold

Sig, significant trend; In-sig, insignificant trend

Table 5 Overall trend analysis for Peninsular Malaysia using MK and Sen' slope tests

Estimate	Weekly		Monthly		NEM season		SWM season	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Z-Value	2.17009	4.59819	1.83799	3.79257	−1.7009	1.03586	2.79298	3.4601
Sen's slope	0.00187	0.0075	0.00019	0.00053	−0.0051	0.00576	0.00919	0.0203
S	35,186	74,546	268,699	554,366	−7424	4521	12,190	15,099
Var(S)	$2.6*10^8$	$2.6*10^8$	$2.1*10^{10}$	$2.1*10^{10}$	$1.9*10^5$	$1.9*10^5$	$1.9*10^5$	$1.9*10^5$
Sig / In-sig	**Sig**	In-sig	**0. Sig**	In-sig	**Sig**	In-sig	**Sig**	**Sig**

Significant values are represented in bold

Table 5 summarizes the overall results of the trend analysis for the whole of Peninsular Malaysia using MK and Sen' slope tests. Based on the results, it can be concluded that Peninsular Malaysia has experienced significant increasing trends in the mean API series in all periods except in the NEM season. A significant decreasing trend in mean API for NEM season was observed. In addition, a significant increasing trend in the maximum API series is only seen in the SWM season.

Innovative trend analysis method for API

Figure 9 shows the results of the innovative trend analysis approach for API series halves with monotonic and non-monatomic trends for the most ten polluted stations during the study period. The IDs of these ten polluted stations are CA0001, CA0006, CA00011, CA0025, CA0041, CA0043, CA0044, CA0056, CA0057 and CA0060, and the time-series halves plots of these stations are shown in Figure 9(a)-(j) respectively. A scatter of points on both sides of the 1 : 1 line (the 45° line) can be clearly observed. The dashed lines above and below the straight line are the 5% and 10% variance lines from the 45° line. It is obvious that the monotone increasing (decreasing) trend in the given time series falls above (below) the 1 : 1 line. Figures 9(a)-(e) and (g) showcase non-monotonic (increasing) trends, where the low (high) values are more (less) in the first half than the second half. Whereas Figures 9(f), (h), and (i) present the opposite situation, where the high (low) values are less (more) in the first half than the second half, indicating that there is an increasing trend in API series during the second half of time series with respect to the first half. These cases correspond to nonmonotonic trends within the same time series there are increasing and decreasing trends at different scales. While in Figure 9(j) almost all data points are in the second half, this time by considering two halves and the sorting procedure, indicating that time series has a monotonic trend; thus, it showcases a monotonically increasing trend. Generally, the mean monthly API time-series for most stations have a composition of various trend patterns (the scatter points take their position on a curve), indicating a nonmonotonic trend. The innovative trend analysis method is easy to conduct, and the cost of the computation is lower when compared to other methods. While in Figure 9(j) almost all data points are in the second half, this time by considering two halves and the sorting procedure, indicating that time series has a monotonic trend; thus, it showcases a monotonically increasing trend. Generally, the mean monthly API time-series for most stations have a composition of various trend patterns (the scatter points take their position on a curve), indicating a non-monotonic trend. The innovative trend analysis method is easy to conduct, and the cost of the computation is lower when compared to other methods. This innovative method does not provide any magnitude of the trend; therefore, it is a qualitative analysis rather than a quantitative analysis.

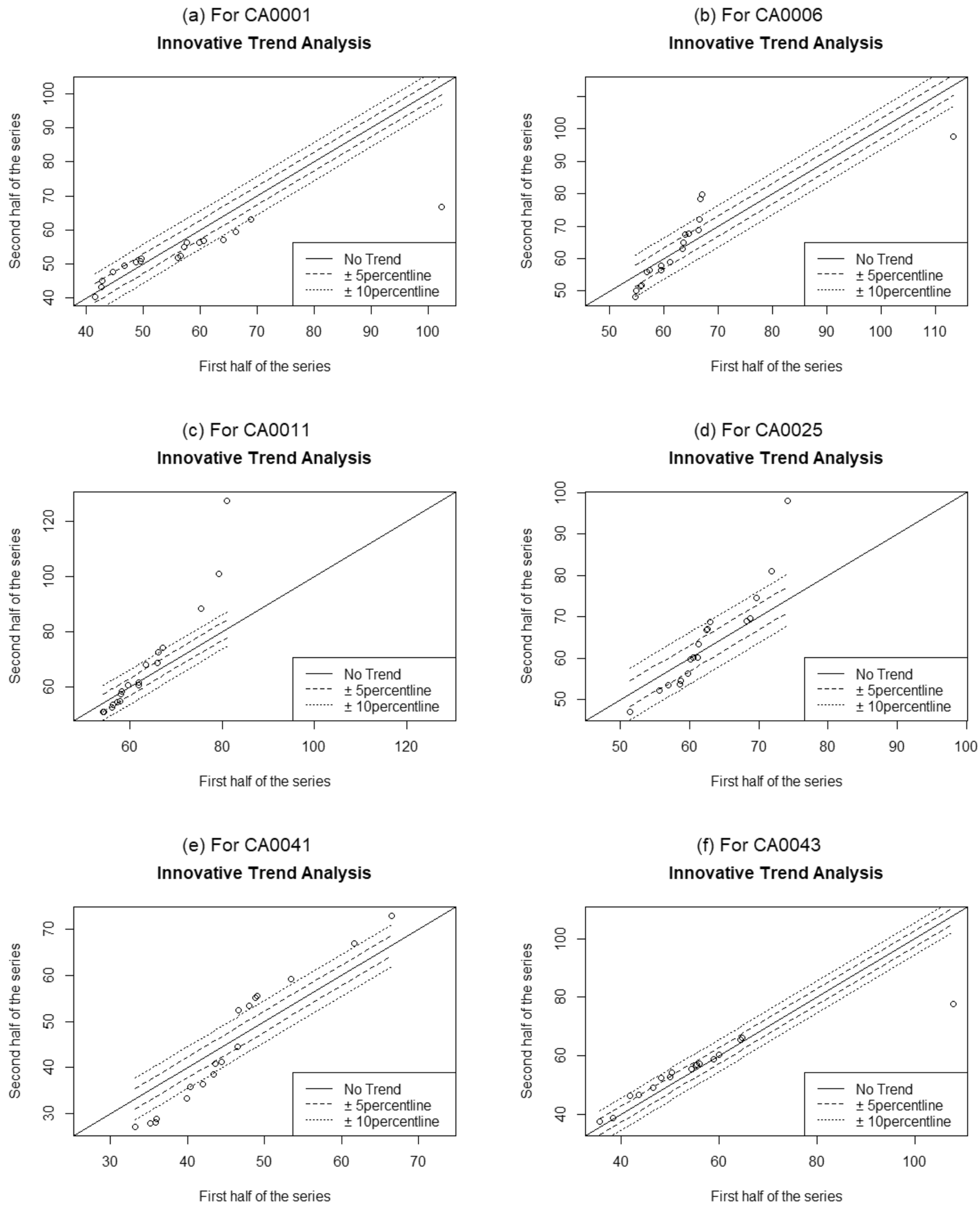

Fig. 9 Mean monthly API time series halves with monotonic and nonmonatomic trends for the most ten polluted stations during the study period (**a**–**j**)

Therefore, this method is well used to identify the possible trends in API series.This innovative method does not provide any magnitude of the trend; therefore, it is a qualitative analysis rather than a quantitative analysis. Therefore, this method is well used to identify the possible trends in API series.

Change-point detection for API

The values of the test statistics and the detection point for the Pettitt test are presented in Table 6. Significant change points exist if the *P* values are below the 5% significance level.

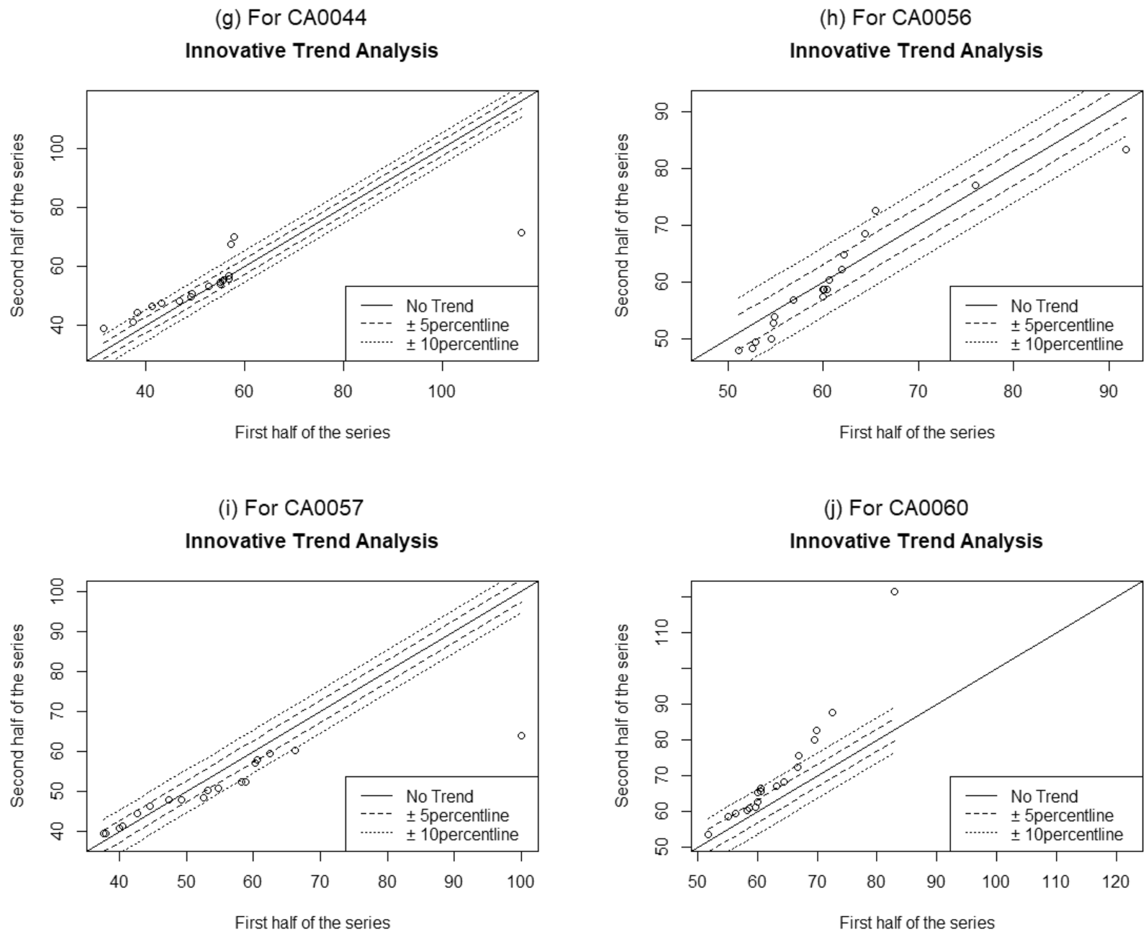

Fig. 9 (continued)

Table 6 Detection of change points via Pettitt test for monthly and seasonal mean API

Station	Monthly API Mean			Southwest Monsoon (SWM)			Northeast Monsoon NEM)		
	U statistic	*P* value	Change point	*U* statistic	*P* value	Change point	*U* statistic	*P* value	Change point
CA0003	169	**0.05**	**January 2014**	24	0.77	No	26	0.65	No
CA0006	167	0.06	No	50	**0.03**	**2013**	42	0.11	No
CA0017	210	**0.01**	**June 2013**	32	0.36	No	50	**0.03**	**2013**
CA0024	199	**0.01**	No	50	**0.03**	**2013**	34	0.29	No
CA0034	182	**0.03**	**October 2013**	22	0.89	No	50	**0.03**	**2013**
CA0038	173	**0.05**	**January 2014**	26	0.89	No	28	0.54	No
CA0040	109	0.45	No	28	0.54	No	20	1.00	No
CA0044	71	1.00	No	50	**0.03**	**2013**	32	0.36	No
CA0047	201	**0.01**	**March 2013**	34	0.29	No	38	0.18	No
CA0054	176	**0.04**	**August 2013**	48	**0.04**	**2013**	16	1.00	No
CA0057	96	0.63	No	50	**0.03**	**2013**	20	1.00	No
CA0059	246	**0.00**	**August 2013**	44	0.08	No	50	**0.03**	**2013**
CA0060	176	**0.04**	**February 2013**	20	1.00	No	42	0.11	No

Significant values are represented in bold

No significant change points were identified in the seasonal means of API at most stations except a few at Bukit Rambai (CA0006), Paka (CA0024), Muar (CA0044), Cheras (CA0054), and Kota Tinggi (CA0057) from the Pettitt test. The changing points were found at these stations during the SWM season in the year 2013. On the other hand, significant change points of API data can be found for Kedah (CA0017), Kuala Terengganu (CA0034), and Tanah Merah (CA0059) stations during the NEM in year 2013.

The change points are detected for nine stations for monthly mean API. These stations are CA0003, CA0027, CA0034, CA0038, CA0044, CA0047, CA0054, CA0059 and CA0060 and the change points were recorded in months January 2014, June 2013, May 2013, September 2013, January 2014, March 2013, August 2013, August 2013, and February 2013, respectively. Generally, almost all the significant change points captured by the Pettitt test are in the year 2013, and only a few in 2014. The possible

Table 7 Mann–Kendall statistics and Pettitt test for **daily** API concentrations in the study period

Station ID	Mann–Kendall (2012–2014)			Pettitt test (2012–2014)		
	Z	Sen's slope	*P* value	U Statistic	*P* value	Change-point
CA0001	−0.42461	0.0000	0.67112	47,874	**0.0000**	10/25/2012
CA0002	−4.11834	−0.00417	**0.00004**	78,160	**0.0000**	10/16/2012
CA0003	3.65538	0.00541	**0.00026**	71,686	**0.0000**	21/01/2014
CA0006	−2.73072	−0.00284	**0.00632**	112,276	**0.0000**	16/03/2014
CA0007	4.26928	0.0048	**0.00002**	67,788	**0.0000**	16/04/2014
CA0008	3.19176	0.00414	**0.00141**	68,062	**0.0000**	23/01/2014
CA0009	1.58756	0.00152	0.11238	40,875	**0.0009**	20/01/2014
CA0010	−0.68405	0.0000	0.49395	40,741	**0.0010**	18/10/2012
CA0011	0.35913	0.0000	0.7195	34,155	**0.0098**	09/06/2012
CA0014	−1.92289	−0.00188	0.05449	46,269	**0.0001**	01/10/2013
CA0015	−1.50731	−0.00115	0.13173	57,033	**0.0000**	12/10/2012
CA0016	2.59697	0.00323	**0.00941**	48,655	**0.0000**	19/01/2014
CA0017	12.83583	0.01412	**0.00000**	139,601	**0.0000**	14/06/2013
CA0019	1.92513	0.00214	0.05421	38,549	**0.0000**	28/03/2014
CA0020	4.79154	0.00654	**0.00000**	107,395	**0.0000**	13/01/2014
CA0022	5.33921	0.00575	**0.00000**	64,033	**0.0000**	13/07/2013
CA0024	8.90076	0.01001	**0.00000**	111,756	**0.0000**	12/06/2013
CA0025	1.30723	0.00197	0.19114	44,876	**0.0002**	19/01/2014
CA0032	−1.42922	0.0000	0.15294	67,434	**0.0000**	26/07/2014
CA0033	2.98421	0.00342	**0.00284**	80,649	**0.0000**	04/01/2013
CA0034	−10.0302	−0.01099	**0.00000**	126,325	**0.0000**	02/10/2013
CA0038	4.04962	0.00615	**0.00005**	60,206	**0.0000**	10/01/2013
CA0040	−0.1191	0.0000	0.90519	55,645	**0.0000**	14/01/2014
CA0041	−1.78494	−0.00224	0.07427	68,792	**0.0000**	31/08/2013
CA0043	1.07351	0.00116	0.28304	59,194	**0.0000**	05/03/2013
CA0044	−0.03909	0.0000	0.96881	49,797	**0.0000**	12/10/2012
CA0045	1.39809	0.00205	0.16209	44,588	**0.0002**	26/10/2014
CA0046	0.78085	0.0000	0.43489	41,773	**0.0007**	06/08/2014
CA0047	4.43039	0.00739	**0.00001**	96,671	**0.0000**	09/03/2013
CA0048	−4.3585	−0.00635	**0.00001**	72,749	**0.0000**	31/08/2013
CA0053	0.26879	0.0000	0.78809	42,291	**0.0006**	21/01/2014
CA0054	−4.87402	−0.00974	**0.00000**	72,758	**0.0000**	05/08/2013
CA0056	−1.87788	−0.00251	**0.06040**	34,819	**0.0080**	20/05/2012
CA0057	−3.73813	−0.0043	**0.00019**	71,253	**0.0000**	29/11/2012
CA0058	3.05542	0.00648	**0.00225**	60,340	**0.0000**	30/01/2014
CA0059	−14.4734	−0.01869	**0.00000**	169,953	**0.0000**	26/09/2013
CA0060	5.24309	0.00815	**0.00000**	67,077	**0.0000**	21/02/2013

Significant values are represented in bold

days for changing points are given in Table 6 and they can be related to the phenomenon of El-Nino.

Likewise, no significant change points are detected by the Pettitt test in the monthly and seasonal maximum API series for most of the stations (these results are shown in Supplementary Table S5). However, several exceptions were found. They are CA0006, CA0024, and CA0047 stations, and the change points in the monthly maximum API series were observed in February 2013, May 2013, and February 2013, respectively. In addition, change points were detected at two stations CA0006 and CA0047 during the NEM season for 2013 and SWM season for 2012, respectively. In general, most of the detected change points were found in the years 2012 and 2013. A relationship might be assumed between the captured detection points to the El-Nino phenomenon. However, further research has to be extensively conducted to confirm such a conclusion.

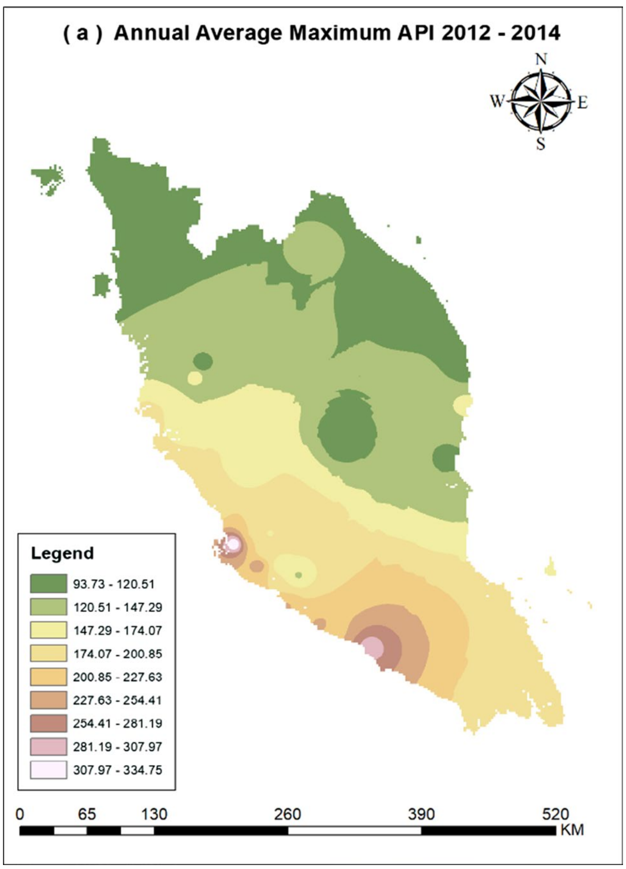

Fig. 10 IDW results for describing the trend of API occurrences in Peninsular Malaysia

Table 7 presents the comparison results of the Mann–Kendall statistic and Pettitt test for daily API concentrations during the study period. The table showcases the increasing and decreasing trends with respect to their changing points. The bold sections are the significant trends.

Spatial distribution map of air pollution

The inverse distance weighted (IDW) interpolation method was employed based on ArcGIS software for investigating the increasing and decreasing trend of API series. This can be easily used to characterize the trends of air pollution events. IDW results describe the trend for air pollution events in Peninsular Malaysia, which interprets an upward trend, especially in southwest parts. The majority of parts of Peninsular Malaysia are expected to experience a moderate status of air quality according to the results in Fig. 10.

Figure 10 further showcases that the observed changes in air pollution are related to the regional effect. This indicates that the regional distributions of API in Peninsular Malaysia over time are spatially clustered.

With all these results, further attention should be given to the high-risk regions such as Johor, Klang, and Melaka for the prevention and mitigation of air pollution effects. Thus, the damage to the public and the environment in the future can be minimized.

Conclusion

Various aspects, including upward trends, downward trends, and changing points of trends related to API on the air pollution to Malaysia are discussed in this study. The trend analysis and change point detections of monthly and seasonal means and maximum API series were carried out for 37 air monitoring stations in Peninsular Malaysia using the nonparametric Mann–Kendall and Pettitt tests. A mix of decreasing and increasing trends can be observed at the API series of the air monitoring stations studied. Furthermore, advantages like the convenience of conduct and lowered computational cost were observed for the innovative trend method compared to the MK method. However, this innovative method is a qualitative analysis rather than a quantitative analysis. Therefore, this method can be used to identify the possible trends in API series. In general, it can be concluded herein that the graphical analysis can be easily adapted for the API trend analysis for the qualitative analysis; however, for quantitative analysis, a statistical trend analysis is highly recommended.

The change point detection via the Pettitt test showed that most of the detection points at the stations in both monthly and seasonal analyses have occurred in the year 2013. Therefore, a general pattern of moderate air pollution state is observed in this study. Both monthly and seasonal means and maximum API are significant for a few stations. However, most of the other stations were found with non-significant increasing trends. The significant monitoring trends in the API series were found to be related to the presence of significant change points that have been detected at the station. In general, the observed status of unhealthy air pollution could have been impacted by the occurrence of the El-Nino phenomenon and other climate factors such as haze, which occurred due to the forest fires from Indonesia.

Many studies have shown that urban areas are more polluted than rural areas due to the presence of heavy traffic, which emits a large number of pollutants into the atmosphere. The urban condition as well as the industrial condition, which are known to be factors that cause air pollution, have not been explicitly taken into account but implicitly taken care of in this study. The urbanized and industrialized effects on API can be observed based on higher values of monthly and seasonal mean and maximum API found in the cities and industrial areas. In addition, the rural areas have experienced lower API values. This study could have been made more interesting with micro-metrological parameters, such as relative humidity, wind speed, wind direction, and air. However, future research would be needed to state solid conclusions. Nevertheless, the impact may be swamped by El-Nino occurrences. Thus, future research is proposed with all governing parameters for a holistic study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13762-021-03672-w.

Acknowledgements The authors are grateful to the Department of Environment Malaysia for their cooperation in providing the air pollution data. The authors also wish to thank the anonymous reviewers for their critical comments and views that led to the improvement of this paper and express thanks to Mr. Abdullah Al-Yaari for reviewing the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Afroz R, Hassan MN, Ibrahim NA (2003) Review of air pollution and health impacts in Malaysia. Environ Res 92(2):71–77

Al-Dhurafi NA, Masseran N, Zamzuri ZH, Razali AM (2018a) Modeling unhealthy air pollution index using a peaks-over-threshold method. Environ Eng Sci 35(2):101–110

Ali R, Kuriqi A, Abubaker S, Kisi O (2019) Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann–Kendall and Sen's innovative trend method. Water 11(9):1855

Alyousifi Y, Ibrahim K, Kang W, Zin WZW (2021) Robust empirical Bayes approach for Markov chain modeling of air pollution index. J Environ Health Sci Eng 19(1):343–356

Alyousifi Y, Ibrahim K, Kang W, Zin WZW (2019) Markov chain modeling for air pollution index based on maximum a posteriori method. Air Qual Atmos Health 12(12):1521–1531

Alyousifi Y, Ibrahim K, Wei K, Zin WZW (2020a) Modeling the spatio-temporal dynamics of air pollution index based on spatial markov chain model. Environ Monit Assess 11(192):1–24

Alyousifi Y, Masseran N, Ibrahim K (2018) Modeling the stochastic dependence of air pollution index data. Stoch Env Res Risk Assess 32(6):1603–1611

Alyousifi Y, Othman M, Sokkalingam R, Faye I, Silva PC (2020b) Predicting daily air pollution index based on fuzzy time series markov chain model. Symmetry 12(2):293–310

Arman M, Azid A, Juahir H, Toriman M, Mustafa A, Hasnam C, Yunus K (2015) Spatial analysis of the certain air pollutants using environmetric techniques. Jurnal Teknologi 75(1):241–249

Azid A, Juahir H, Aris AZ, Toriman ME, Latif MT, Zain SM, Saudi ASM (2014) Spatial analysis of the air pollutant index in the Southern Region of Peninsular Malaysia using environmetric techniques. From sources to solution. Springer, pp 307–312

Azmi SZ, Lati MT, Ismail AS, Juneng L, Jemain AA (2010) Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual Atmos Heal 9(3):53–64

AL-Dhurafi N A, Masseran N, Zamzuri Z H, (2020) Hierarchical-Generalized Pareto Model for Estimation ofUnhealthy Air Pollution Index. Environ Model Assess 15(5):1–10
Butler CD, Whelan J (2018) Air pollution and climate change in Australia: a triple burden. Climate change and air pollution. Springer, pp 131–149
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. Environ Monit Assess 186(8):4719–4742
Dabanlı İ, Şen Z, Yeleğen MÖ, Şişman E, Selek B, Güçlü YS (2016) Trend assessment by the innovative-Şen method. Water resources management 30(14):5193–5203
Damberg L, AghaKouchak A (2014) Global trends and patterns of drought from space. Theoret Appl Climatol 117(3–4):441–448
Demir V, Kisi O (2016) Comparison of Mann-Kendall and innovative trend method (Şen trend) for monthly total precipitation. Middle Black Sea Region, Turkey
Deni SM, Suhaila J, Wan Zin WZ, Jemain AA (2010) Spatial trends of dry spells over Peninsular Malaysia during monsoon seasons. Theor Appl Climatol 99:357–371
Department of Environment Malaysia (DOE) (2000) A guide to Air Pollutant Index in Malaysia (API). Ministry of Science, Technology, and the Environment, Kuala Lumpur, Malaysia
Dominick D, Juahir H, Latif MT, Zain SM, Aris AZ (2012) Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmos Environ 12(60):172–181
Douglas EM, Vogel RM, Kroll CN (2000) Trends in floods and low flows in the United States: impact of spatial correlation. J Hydrol 240(4):90–105
Gass K, Klein M, Sarnat S, Winquist A, Darrow L, Flanders W, Chang H, Mulholland J, Tolbert P, Strickland M (2015) Associations between ambient air pollutant mixtures and pediatric asthma emergency department visits in three cities: a classification and regression tree approach. Environ Health 7(58):1–14
Halim NDA, Latif MT, Ahamad F, Dominick D, Chung JX, Juneng L, Khan MF (2018) The long-term assessment of air quality on an island in Malaysia. Heliyon 4(12):10–54
Hirsch R, Slack J, Smith R (1982) Techniques of trend analysis for monthly water quality data. Wat Res Res 32(1):107–121
Ismail M, Suroto A, Ismail NA (2012) Time series analysis of surface ozone monitoring records in Kemaman. Malaysia. Sains Malaysiana 40(5):411–417
Jaiswal A, Samuel C, Kadabgaon VM (2018) Statistical trend analysis and forecast modeling of air pollutants. Glob J Environ Sci Manag 4(4):427–438
Jaiswal RK, Lohani AK, Tiwari HL (2015) Statistical analysis for change detection and trend assessment in climatological parameters. Environ Process 2(4):729–749
Kendall M G (1948) The advanced theory of statistics, 4th edn, vol 1
Kendall MG (1975) Rank correlation methods. Griffin
Latif MT, Azmi SZ, Noor AD, Ismail AS, Johny Z, Idrus S, Mohamed AF, Mokhtar M (2011) The impact of urban growth on regional air quality surrounding the Langat River Basin. Malaysia. Environmentalist 31(12):315–324
Maleki H, Sorooshian A, Goudarzi G, Baboli Z, Birgani YT, Rahmati M (2019) Air pollution prediction by using an artificial neural network model. Clean Technol Environ Policy 21(6):1341–1352
Mann B (1945) Nonparametric tests against trend. Econom J Econom Soc 13:245–259
Millard S P, Kowarik A, Kowarik M A (2020) Package 'EnvStats'. version 2.2.1 http://CRAN.R-project.org
Mitchell W J (1994) The reconfigured eye: Visual truth in the post-photographic era. Mit Press
Militino AF, Moradi M, Ugarte MD (2020) On the performances of trend and change-point detection methods for remote sensing data. Remot Sens 12(6):1008
Mohtar AAA, Latif MT, Baharudin NH, Ahamad F, Chung JX, Othman M, Juneng L (2018) Variation of major air pollutants in different seasonal conditions in an urban environment in Malaysia. Geosci Lett 5(1):1–13
Motiva MI, Bielik N, Campelos PB, Cluse C, Goujon-Ginglinger CG, Jaquier A, Polier-Calame J (2019) Air quality assessment of the Tobacco Heating System 22 under simulated residential conditions. Air Qual Atmos Health 12:807–883
Nieto PG, Lasheras FS, García-Gonzalo E, de Cos Juez FJ (2018) PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study. Sci Total Environ 621:753–761
Öztopal A, Şen Z (2017) Innovative trend methodology applications to precipitation records in Turkey. Wat Resourc Manag 31(3):727–737
Pal I, Al-Tabbaa A (2009) Trends in seasonal precipitation extremes—an indicator of 'climate change'in Kerala, India. J Hydrol 367(1–2):62–69
Patakamuri S K, O'Brien N, Patakamuri M S K (2020) Package 'modifiedmk'. Version 1.5.0. CRAN
Pettitt AN (1979) A non-parametric approach in the change point problem. Appl Stat 28(2):126–135
Pohlert T, Pohlert MT, Kendall S (2016) Package 'trend'. Title Non-Parametric Trend Tests and Change-Point Detection
Rahman SRA, Ismail SNS, Raml MF, Latif MT, Abidin EZ, Praveena SM (2015) The assessment of ambient air pollution trend in Klang Valley, Malaysia. World Environ 5(1):1–11
Rathnayake U (2019) Comparison of statistical methods to graphical methods in rainfall trend analysis—case studies from tropical catchments. Adv Meteorol 2019:1–10
Abd Rani NL, Azid A, Khalit SI, Juahir H, Samsudin MS (2018) Air pollution index trend analysis in Malaysia, 2010–15. Pol J Environ Stud 27(2):801–807
Salarijazi M, Akhond-Ali A, Daneshkhah Adib A, A, (2012) Trend and changepoint detection for the annual streamflowseries of the Karun River at the Ahvaz hydrometric station. African J Agricult Res 7(32):4540–4552
Sen A, Srivastava M (2012) Regression analysis: theory, methods, and applications. Springer
Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. J Am Stat Assoc 63(324):1379–1389
Şen Z (2017) Innovative trend significance test and applications. Theoret Appl Climatol 127(4):939–947

Singh KP, Bartolucci AA, Bae S (2001) Mathematical modeling of environmental data. Mathemat Computer Model 33(6–7):793–800

Suhaila J, Deni SM, Wan Zin WZ, Jemain AA (2010) Spatial patterns and trends of daily rainfall regime in Peninsular Malaysia during the Southwest and Northeast Monsoons: 1975–2004. Meteorol Atmos Phys 110:1–18

Suhaila J, Yusop Z (2018) Trend analysis and change point detection of annual and seasonal temperature series in Peninsular Malaysia. Meteorol Atmos Phys 130(5):565–581

Tabari H, Taye MT, Willems P (2015) Statistical assessment of precipitation trends in the upper Blue Nile Riverbasin. Stochast Environ Res Risk Assess 29(7):1751–1761

Theil H (1950) A rank-invariant method of linear and polynominal regression analysis (Parts 1–3). In Ned. Akad Wetensch Proc Ser A 53:1397–1412

Tosunoglu F, Kisi O (2017) Trend analysis of maximum hydrologic drought variables using Mann–Kendall and Sen's innovative trend method. River Res Appl 33:597–610

Von Storch H (1999) Misuses of statistical analysis in climate research. Analysis of climate variability. Springer, pp 11–26

Wang S, Zhang X (2012) Long-term trend analysis for temperature in the Jinsha River Basin in China. Theoret Appl Climatol 109(3):591–603

Yusof F, Hui Mean F, Suhaila J, Yusop Z (2013) Characterisation of drought properties with bivariate copula analysis. Water Resour Manag 27:4183–4207. https://doi.org/10.1007/s11269-013-0402-4

Zarenistanak M, Dhorde AG, Kripalani RH (2014) Trend analysis and change point detection of annual and seasonal precipitation and temperature series over southwest Iran. J Earth Syst Sci 123(2):281–295

Zhang Q, Streets DG, Carmichael GR, He K, Huo H, Kannari A, Klimont Z, Park I, Reddy S, Fu J (2009) Asian emissions in 2006 for the Nasa Intex-B Mission. Atmos Chem Phys 9(14):5131–5153

Zhang L, Lin J, Qiu R, Hu X, Zhang H, Chen Q, Wang J (2018) Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model. Ecol Ind 95:702–710