Assessing the impacts of climate variables on long-term air quality trends in Peninsular Malaysia.

ZHENG, Y., OOI., M.C.G., JUNENG, L., WEE, H.B., LATIF, M.T., NADZIR, M.S.M., HANIF, N.M., CHAN, A., LI, L., AHMAD, N.B., TANGANG, F.

2023

Supplementary materials are appended after the main text of this document.

Assessing the impacts of climate variables on long-term air quality trends in Peninsular Malaysia

Yijing **Zheng**[1,2], Maggie Chel Gee **Ooi**[1]*, Liew **Juneng**[3], Hin Boo **Wee**[1], Mohd Talib **Latif**[3], Mohd Shahrul Mohd **Nadzir**[3], Norfazrin Mohd **Hanif**[3], Andy **Chan**[4], Li **Li**[5,6], Norfazilah binti **Ahmad**[7] Fredolin **Tangang**[3]

[1]Insitute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia

[2]School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, People's Republic of China

[3]Department of Earth Sciences and Environment, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, UKM Bangi 43600, Selangor, Malaysia

[4]Department of Civil Engineering, University of Nottingham Malaysia, Semenyih 43500, Selangor, Malaysia

[5]School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China

[6]Key Laboratory of Organic Compound Pollution Control Engineering (MOE), Shanghai University, Shanghai, China

[7] Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia 56000 Cheras, Kuala Lumpur Malaysia

[*] Corresponding author: chelgee.ooi@ukm.edu.my

ABSTRACT

Climate change is thought to influence the composition of atmospheric air, but little is known about the direct relationship between these variables, especially in a hot tropical climate like that of Malaysia. This work summarizes and analyzes the climate state and air quality of Peninsular Malaysia based on selected ground-based observations of the temperature, precipitation, relative humidity, wind speed, wind direction and concentrations of PM_{10}, O_3, CO, NO_2, and SO_2 over the last 20 years (2000–2019). The relationship between the climate state and air quality is analyzed using the Pearson correlation and canonical correlation analysis (CCA) methods is employed to predict the degree of change in the future air quality under different warming scenarios. It is found that the Peninsular Malaysia mainly experienced strong precipitation in the central and mountainous regions, while air pollutants are primarily concentrated in densely populated areas. Throughout the period of study (interannual, monthly, and diurnal time series analyses), Peninsular Malaysia became warmer and drier, with a significant increase in temperature (+4.2%), decrease in the relative humidity (-4.5%), and greater fluctuation in precipitation amount. The pollution conditions have worsened; there has been an increase in the PM_{10} (+16.4%), O_3 (+39.5%), and NO_2 (+2.1%) concentration over the last 20 years. However, the amount of SO_2 (-53.6%) and CO (-20.6%) decreased significantly. The analysis of the monthly variation shows a strong bimodality of the PM_{10} and O_3 concentrations that corresponds to the monsoon transition. Intensive diurnal fluctuations and correlations are observed for all the variables in this study. According to the CCA, the air quality factors are strongly correlated with meteorological factors; in particular, the CO, O_3, and PM_{10} concentrations interact strongly with

the air temperature. These findings show that the future air quality in Peninsular Malaysia has high possibility to deteriorate under warming condition.

Keywords: Global warming, temperature, ozone, canonical correlation analysis

1 INTRODUCTION

World Health Organization (WHO) reports indicate that outdoor air pollution was estimated to have caused 4.2 million premature deaths worldwide in 2016, and 91% of these early deaths happened in low- and middle-income countries, with the majority occurring in Southeast Asia and the Western Pacific (WHO, 2021). Climate change is projected to further lower the air quality in polluted areas because of negative changes in air pollution meteorological processes (IPCC, 2014). The Intergovernmental Panel on Climate Change (IPCC) has reported that global average temperatures have increased by about 1.1°C since pre-industrial times and that most of the warming observed over the past 50 years can be attributed to human activities (IPCC, 2021). A rise in temperature can change the dynamics of the air composition in several ways. One key mechanism is the increase in the chemical reactions in the atmosphere that occur at higher temperatures, which can lead to increased levels of air pollutants, including ozone and particulate matter (Seinfeld and Pandis, 2016). Another important mechanism is the release of greenhouse gases from various sources, including permafrost, wetlands, and the ocean (IPCC, 2021; Schuur et al., 2015).

Under the background of obvious global climate change, there is a growing interest in studying how climate change and its consequences for air quality play out locally, particularly in Malaysia, which experiences a tropical wet climate (Tang, 2019). Malaysia is a Southeast Asian country divided into two sections by the South China Sea: Peninsular Malaysia and Borneo's East Malaysia. Peninsular Malaysia, which accounts for 40% of Malaysia's total land area, is where the major cities are concentrated. The climates of Peninsular and East Malaysia diverge because the climate of Peninsular Malaysia is directly impacted by the wind from continental Asia, while the climate in East Malaysia is much more affected by marine circulation. There are two monsoonal seasons due to the seasonal variation of the Intertropical Convergence Zone (ITCZ) and the related trade wind fields in the area (Sentian et al., 2019). The southwest and northeast monsoons have a strong influence on the yearly climatic variability. The southwest monsoon lasts from April to September, whereas the northeast monsoon lasts from October to March. In comparison to the northeast monsoon, which provides greater precipitation, the southwest monsoon has drier climate and less rainfall (Kwan et al., 2013). Two prominent interannual signals—the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD)—and an intra-seasonal signal—the Madden-Julian Oscillation (MJO)—can cause significant variations in Malaysia's climate and impact the air quality in the region, leading to droughts, floods, and other climate-related hazards, with increased air pollution levels during active ENSO, IOD, and MJO phases (Islam et al., 2018; Jud et al., 2020; Kuwata et al., 2021; Tangang et al., 2017; Xiao et al., 2022). The impact of climate on air pollution complicates air quality management efforts, especially for a developing country like Malaysia. Accurately understanding, mastering, and forecasting the patterns and characteristics of climate and air quality conditions are of great

significance for disaster prevention, mitigation, adaptation, and the scientific planning of production.

From 2000 to 2019, according to Malaysia's Fourth Biennial Update Report (BUR 4), greenhouse gas (GHG) emissions have increased substantially for about 7 times. Excluding Land-use Change and Forestry (LULUCF), the energy sector was the leading source of emissions, accounting for an average of 80% of annual emissions from 2000 to 2019. Numerous studies have shown that the emission of GHG warms the atmosphere and has a significant positive correlation with temperature (Andrée et al., 2019; Mikhaylov et al., 2020; Neagu & Teodoru, 2019). According to studies conducted in Peninsular Malaysia, a significant warming trend has been observed in recent decades (Ling, 2009; Suhaila and Yusop, 2018; Tangang et al., 2006; Wong, 2018). This rise in temperature has led to an increase in the frequency and intensity of precipitation events and more frequent floods (Mayowa et al., 2015; Ng et al., 2022). Many scientists have also studied and analyzed the air quality in Malaysia. Morrissey et al. (2021) found that, while the air quality in the Greater Kuala Lumpur region is improving, no level of air pollution can be considered acceptable. Long-term (1997–2015) PM_{10} pollution in Malaysia is decreasing at a slow but considerable rate (Sentian et al., 2019). Significant increases in ozone concentrations were reported in Malaysia (Ahamad et al., 2020; Ismail et al., 2011). Malaysia has witnessed significant warming, rainfall anomalies, and a significant upward trend for ozone over the last two decades, attracting considerable interest in the study of climate trends and their consequences, especially for air quality (Dominick et al., 2012; Halim et al., 2018; Malaysia, 2009; Suhaila and Yusop, 2018; Tang, 2019). Empirical orthogonal function (EOF) analysis has been used in several studies to investigate the spatiotemporal patterns and variability of

meteorological and air quality variables in the region (Juneng et al., 2009; Khoir et al., 2022). Quantitative knowledge of how air pollution reacts to both global warming and variability at the regional scale could inform air quality planning in the future.

This study aims to assess the relationship between long-term climate change and air quality conditions in Peninsular Malaysia and predict the future air quality under different temperature-rise scenarios. The first part of the study analyzes the climate and air quality in Peninsular Malaysia (henceforth referred to as the Peninsula) from 2000 to 2019. With this information, this study identifies the long-term relationships between climate factors and the variation of air quality variables through Pearson correlation analysis and canonical correlation analysis (CCA). Finally, this study predicts the future air quality under different temperature-rise scenarios using the machine learning-based CCA algorithm. By analyzing historical data and projecting future trends, this study seeks to provide insights into the potential impacts of climate change on the air quality in the region and inform policy decisions aimed at mitigating these impacts.

2 DATA AND METHODOLOGY

2.1 Site and data description

The observational data are obtained from observatories in the Peninsular Malaysia, and they include the hourly temperature, relative humidity, wind speed, wind direction, the concentrations of particulate matter (PM_{10}), ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2), and daily mean rainfall, resulting in a total of 10 variables, ranging from

01/01/2000 to 12/31/2019. The study for year 2020 is excluded due to the inconsistencies of air quality and weather pattern due to short-term interference from the COVID-19 condition (see Supplementary document: S6). The rainfall data are provided by 25 weather stations from the Malaysian Meteorological Department (MMD). The wind data are provided by 10 general weather stations of MET Malaysia from the Global Hourly - Integrated Surface Database (ISD). Other data are provided by the continuous air quality monitoring system (CAQMS) installed by the Department of the Environment (DOE). The data (except for the rainfall and wind data) from 2000 to 2016 are supplied by the Alam Sekitar Malaysia Sdn Bhd (ASMA) contractor. The data from 2017–2019 are provided by the Pakar Scieno TW Sdn Bhd (PSTW) supplier (the new contractor of the DOE). The selection of the 25 weather stations and 10 general stations was made according to the availability of long-term rainfall/wind data and to obtain a geographically even distribution across the Peninsula. Based on the principle of maximizing the use of data, 30 DOE stations (for temperature, relative humidity, and PM_{10}) and 25 other stations (for O_3, CO, NO_2, and SO_2) were chosen because they had 20 years of continuous air quality data and maintain positional consistency before and after 2017. Due to the characteristics of the data studied, which had a wide time range and multiple sites, in order to ensure the validity and representativeness of the data, only sites for which missing values make up less than 20% of the data are selected. The spatial distribution of these stations is given in Figure 1, and information on the data availability is presented in Table 1. By analyzing these variables, the past characteristics of climate change and past air quality conditions in the Peninsula can be fully revealed.

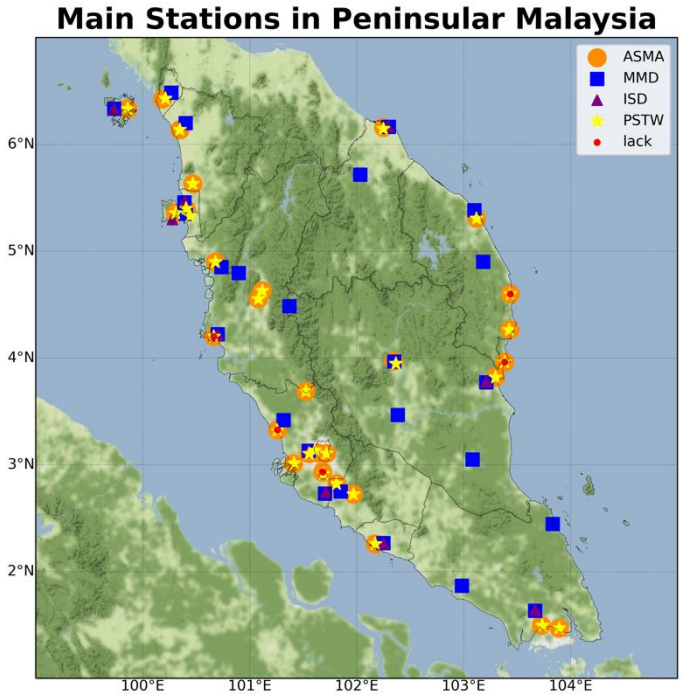


Fig. 1 Locations of the data used in this research, including continuous air quality monitoring system (CAQMS) stations from PSTW (yellow stars) and ASMA (orange circles), general stations for wind data from the ISD (purple triangles), and Malaysian meteorological department weather stations (blue squares). Note that stations in red circles lack air quality data concerning the concentrations of O_3, CO, NO_2, and SO_2.

Table 1

Data availability details for the climate and air quality variables.

Variable	Temperature	Relative humidity	PM_{10}	O_3	CO	NO_2	SO_2	Wind speed	Wind direction	Rainfall
Supplier	ASMA/PSTW							ISD		MMD
Station markers (Caption of Figure 1)	Orange circles / Yellow stars (Stations in red circles lack air quality data concerning the concentrations of O_3, CO, NO_2, and SO_2)							Purple triangles		Blue squares
Period	2000–2016/2017–2019							2000–2019		2000–2019

Note: The full names and abbreviations of the air quality variables in the table are particulate matter (PM_{10}), ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2).

2.2 Meteorological and air quality datasets

During the monitoring periods, all air quality monitoring stations were outfitted with continuous automated monitoring technology designed to collect and measure data continually. The schedules for calibration and upkeep included daily automatic calibration for all contaminants and monthly repairs. The data provided by the DOE for use here have undergone calibration and maintenance regimens devised in accordance with US Environmental Protection Agency standards (Ahamad, 2020). During the process of data provider conversion, although the measurement data went through a quality assurance and quality control process to ensure the accuracy and quality of the data, alterations in instrumentation and workflows may have resulted in discontinuities in the data before and after 2017. In the supplementary information (S1), the original data are displayed on a timing plot, and it can be clearly seen from this figure that there were cliff-like changes in the original data in 2017. To make up for the inconsistency of the data, this study combines and compares the data from ASMA and PSTW to CAMS global reanalysis (EAC4) data from 2017 (Inness et al., 2019), and the PSTW data are calibrated from 2017 to 2019. The EAC4 dataset was chosen because this dataset contains all the parameters required for this study and has surface concentration data that can be directly compared with ground-based observations. The detailed calibration and adjustment process is presented in the supplementary information (S1). The calibrated data have better continuity with the ASMA data. The following research is based on the data after calibration.

2.3 Methodology

The air pollutant concentrations and variations of the meteorological variables were analyzed on annual, monthly, and diurnal scales. For precipitation, the daily observations were added together to provide monthly and yearly precipitation datasets. For other factors, hourly data were used for the computation of the diurnal variations, while monthly and yearly averaged data from hourly readings were used for the monthly and annual variations. The Pearson correlation was calculated across each variable to identify the level of correlation on yearly, monthly, and hourly scales. This allows a more detailed understanding of the correlation of one or more attributes with other attributes at different time scales, and it lays the foundation for subsequent research.

Subsequently, the CCA procedure was used in the present study to reveal the degree of association between climate change factors and air quality factors. By increasing the Pearson correlation between linear combinations of two sets of variables, CCA provides a generic multivariate approach for studying correlations when both sets of variables are quantitative (Akbaş, 2005; Dattalo, 2014; Langworthy et al., 2021). Unlike other methods, it has the ability to linearly connect two distinct variables regardless of their units (Zhang et al., 2020). These linear projections may be thought of as reflecting elements of the data's structure and may thus be beneficial for downstream prediction tasks. Although it is an extensively used statistical technique in many fields like social science, medical research, psychological research, and marketing analytics (Wang et al., 2020; Yang et al., 2019; Zhuang et al., 2020), it is seldom employed in atmospheric science which is a good attempt in the paper (Bowo et al., 2020; Rana et al., 2018; Zhang et al., 2020). Since the dimensionality of the dataset is relatively low, it is feasible to use CCA directly to identify the relationships between the climate data and air quality

data, without the need for dimensionality reduction using EOF analysis. The model theory and calculation method of CCA are as follows.

Two sets of variables X and Y are considered, where X is a set of meteorological variables that includes temperature, relative humidity, wind speed, wind direction, and rainfall data, while Y is a set of air quality variables that includes the concentrations of PM_{10}, O_3, CO, NO_2, and SO_2:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{pmatrix}, p \le q. \tag{1}$$

With these sets of variables, two sets of linear relations U and V are defined, where U contains the linear combinations of X, and V contains the linear combinations of Y:

$$\begin{aligned} U_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p, \\ U_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p, \\ &\vdots \\ U_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p, \end{aligned} \tag{2}$$

$$\begin{aligned} V_1 &= b_{11}Y_1 + b_{12}Y_2 + \cdots + b_{1q}Y_q, \\ V_2 &= b_{21}Y_1 + b_{22}Y_2 + \cdots + b_{2q}Y_q, \\ &\vdots \\ V_p &= b_{p1}Y_1 + b_{p2}Y_2 + \cdots + b_{pq}Y_q. \end{aligned} \tag{3}$$

In order to find the linear combination that maximizes the correlation in each pair of U_i and V_j, the variance of U_i and V_j are defined as follows:

$$var(U_i) = \sum_{k=1}^{p} \sum_{l=1}^{p} a_{ik}a_{il}cov(X_k, X_l), \tag{4}$$

$$var(V_j) = \sum_{k=1}^{p} \sum_{l=1}^{q} b_{jk}b_{jl}cov(Y_k, Y_l) \tag{5}$$

The covariance of U_i and V_j is then calculated as follows:

$$cov(U_i, V_j) = \sum_{k=1}^{p} \sum_{l=1}^{q} a_{ik}b_{jl}cov(X_k, Y_l) \tag{6}$$

The following method is used to evaluate the obtained CCA correlation of U_i and V_j:

$$\rho = \frac{cov(U_i, V_j)}{\sqrt{var(U_i)var(V_j)}}. \quad (7)$$

Hence, to maximize the correlation (ρ_i^*), a linear combination of X and Y that maximizes the abovementioned relationship is determined:

$$\rho_i^* = \frac{cov(U_i, V_i)}{\sqrt{var(U_i)var(V_i)}}. \quad (8)$$

There are generally two methods that can be used for this function optimization. The first is singular value decomposition (SVD), and the second is eigendecomposition. The results obtained by both methods are the same for the dataset, so the output of either of the methods can be used. With the CCA, the first mode of linear correlation is used to determine the variation in the future air quality due to abrupt changes in temperature following future global warming scenarios.

In this study, the Pearson correlation and CCA ($p < 0.001$) computing process is carried out using the Scientific Platform Serving for Statistics Professional (SPSSPRO version 1.0.11) platform, which is a free online application (https://www.spsspro.com). The CCA prediction model is applied in Python with the scikit-learn library, a free and open-source Python machine learning package. It is a simple and effective tool for analyzing predictive data (Pedregosa et al., 2011).

3 RESULTS AND DISCUSSION

3.1 Interannual variation of climate and air quality

Figure 2 shows the interannual variation of the annual average temperature, precipitation, relative humidity, wind speed, PM_{10}, O_3, CO, NO_2, and SO_2 from 2000 to 2019 in the Peninsula. The overall yearly mean temperature showed a steady upward trend, from 26.96°C in 2000 to 28.10°C in 2019, which is consistent with the trend of global warming. Strong fluctuation is observed in the amount of precipitation, and it can be clearly seen that the amplitude of the fluctuation is increasing over time. Although the rainfall amount seems to decrease, the fluctuation amplitude increases, indicating the intensification of extreme precipitation. The humidity decreases in the early 2000s, becomes stable, and then continues to decrease rapidly. The wind speed has a small interannual variation fluctuation of 2–4 years, and the overall change is small, with no significant trend. Long-term changes in climate factors such as the increase in temperature, the increase in the frequency of abnormally warm climate, and the slight decrease in precipitation have been recorded in previous studies (Hanif et al., 2022; Yatim et al., 2019).

CO and NO_2 have similar variation patterns; they increase in the early 2000s, hold relatively constant during the middle 10 years of the considered time period, and then decrease sharply. The concentration of PM_{10} rose until it peaked in 2006 (50.24 $\mu g/m^3$), fell to a second small peak around 2015, and then rose rapidly. O_3 seems to have an increasing trend like that of the temperature over the whole observation period, from its minimum value (15.25 ppb) in 2000 to its maximum value (21.28 ppb) in 2019; this is also consistent with the findings of Ahamad et al. (2020). Surprisingly, SO_2 is one of the few air pollutants that has decreased dramatically over the past 20 years (-2.94 ppb). This is possibly an unanticipated result of clean air measures, including the replacement of polluting coal-fired power plants with clean energy, as well as decreases in the sulfur content in gasoline and diesel, as other comparable studies have demonstrated (Cowern, 2018; Mohtar et al., 2018). The interannual variation of the air quality

factors is somewhat consistent with other studies in the Peninsula, although other studies have focused more on local, short-term data (Halim et al., 2018; Suris et al., 2022).

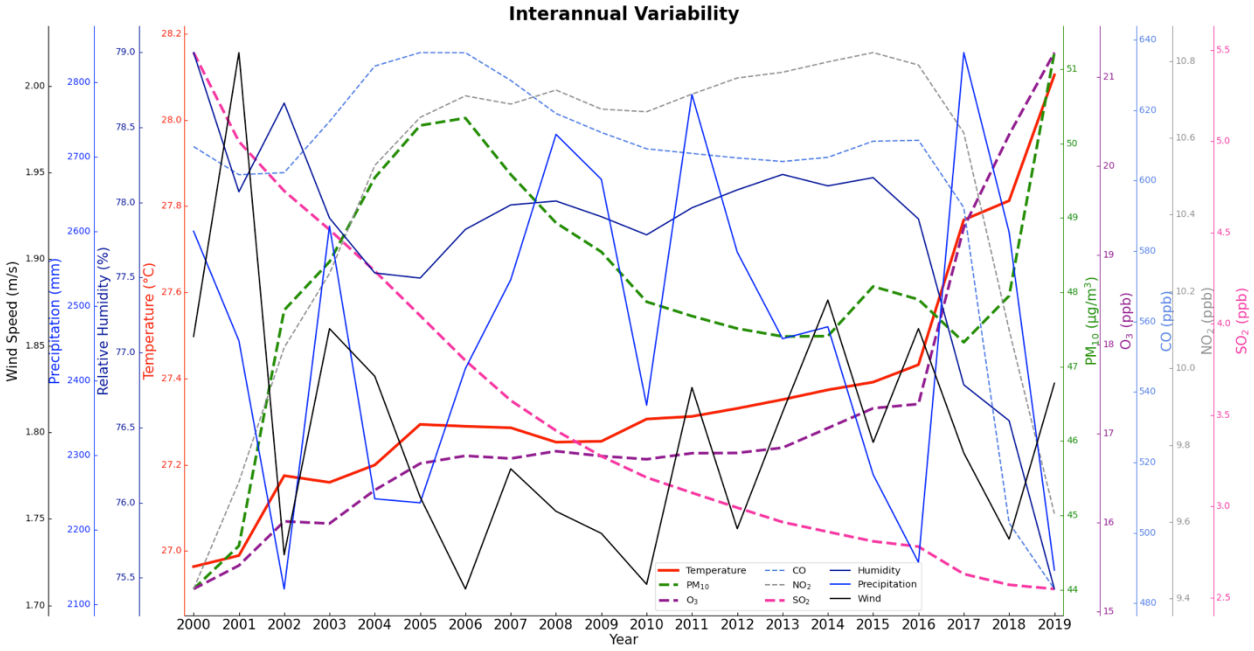


Fig. 2 Interannual variability of yearly climate (mean temperature, relative humidity, precipitation, and wind speed; solid lines) and air quality (PM_{10}, O_3, CO, NO_2, and SO_2; dotted lines) over the observation period.

Table 2

Change in annual mean climate and air quality parameters between 2000 and 2019 over the entire Peninsula.

Period (2000–2019)	Temperature (ºC)	Relative humidity (%)	Rainfall (mm)	PM_{10} (µg/m^3)	Wind speed (m/s)	O_3 (ppb)	CO (ppb)	NO_2 (ppb)	SO_2 (ppb)
Mean (standard deviation)	27 (0.3)	78 (0.8)	2464 (217)	48 (1.7)	1.8 (0.1)	17 (1.5)	602 (39)	10 (0.4)	4 (0.9)
Difference	1.1	-3.6	-454	7.2	-0.03	6	-125.6	0.2	-2.9
Percentage change from 2000	+4%	-5%	-18%	+16%	-2%	+40%	-20%	2%	-54%

Table 2 summarizes the overall changes in all variables between 2000 and 2019 to present a general trend. Macroscopically, temperature, PM_{10}, O_3, and NO_2 levels all increased, with the largest increase in O_3 (+39.49%), followed by PM_{10} (+16.4%), compared to the 2000 values. The relative humidity, precipitation, CO content, and SO_2 content showed a decline, and SO_2 fell the most; it dropped by 53.57% compared with the value from 2000, followed by CO, which fell by 20.6%. Among all the parameters, the wind speed changes the least, with a change of only -1.46%. From the perspective of the standard deviation, the degree of dispersion of precipitation (216.78) is the largest, followed by CO (38.96), PM_{10} (1.72), and O_3 (1.5). Among the air pollutants studied here, the sharp drop in the SO_2 concentration should be noted. In the Peninsula, the main sources of SO_2 emissions are power stations, ships, etc. (Mohtar et al., 2018). The reason for the decline in SO_2 is mainly the government's energy emission control and clean air policy requirements, so it is reasonable to believe that the decline in SO_2 is mainly due to human intervention rather than climate change itself. Although PM_{10} had a slowly declining curve in the early 21st century, it has shown a clear upward trend in volatility in recent years. NO_2 rose sharply after 2000, and gradually slowed down in 2006, showing a flat period. Although it has declined slightly recently, the overall trend is upward. Ozone has been on an upward trend for 20 years; recent studies usually predict a negative impact of climate change on the ozone air quality (Fu and Tian, 2019). It is therefore worth mentioning that the emission control policy has been effective for SO_2 but not for PM_{10}, NO_2, and O_3 over the past 20 years.

3.2 Heat map analysis of climate and air quality in the Peninsula

To indicate the overall features of the variables in the Peninsula, Figure 3 shows the spatial distributions of the yearly mean temperature, precipitation, relative humidity, and concentrations of PM10, O3, CO, NO2, and SO2 over the observation period. Heatmaps are drawn from north

(top) to south (bottom) based on site locations. For specific station names and the corresponding locations, please refer to Figure S4.

From Figure 3, it is obvious that the areas with large populations are prone to higher average temperatures, especially in the central-south regions. The mean temperature on the western side of the Peninsula is slightly higher than that on the eastern side. As time passes, we can clearly see a significant increase in the average temperature for most of the stations. The relative humidity has a different pattern. It can be seen from the figure that the places with high humidity are mainly concentrated in the central Peninsula and near the mountainous areas. The humidity is generally maintained at a high level; it is above 70% at most of the stations. The fluctuation of rainfall and the precipitation frequency can be seen throughout the whole Peninsula, and the average annual precipitation amount in mountainous areas is relatively large. Heavy rainfall occurs mainly around Taiping. The number of days with precipitation has a significant positive correlation with the amount of precipitation. This pattern can also be explained by the interactions between the topography and wind directions, land-sea interactions, and atmospheric circulation (Marzuki et al., 2021; Svensson et al., 2002). Overall, the data indicate a humid tropical rainforest climate in Peninsular Malaysia.

Figures 4 and 5 present the overall 20-year wind speed and direction in Peninsular Malaysia. From the wind speed heat map, it can be concluded that the average wind speed of the entire peninsula is very low, and the fluctuation range is between 1 m/s and 2.7 m/s. The average wind speed in the north is larger than that in the south, and the change in the wind speed with time is not obvious.

This is consistent with previous findings that indicate that Malaysia is in a low-wind zone (Hanoon et al., 2022). From the wind rose diagram, the northeasterly wind generally controls the entire peninsula, the intensity of the northeast monsoon is greater than that of the southwest monsoon, and the wind speed in large cities tends to be lower.

Long-term air pollution in Malaysia is characterized mostly by local emissions and transboundary pollution, particularly in metropolitan areas (Sentian et al., 2019). Along the western side of the Peninsula, the average concentration of PM_{10} is slightly larger than in the eastern part. The concentration of PM_{10} fluctuates severely, mainly from 30–75 $\mu g/m^3$. Large PM_{10} concentrations occur over the western sections where the big cities are located. This is consistent with the connection between larger populations and pollution. From 2002–2006 and during 2015, PM_{10} increased significantly. The high observed value in 2015 is mainly due to the fact that 2015 was an extreme ENSO year, and it was highly correlated with intensive biomass burning episodes in Southeast Asia (Islam et al., 2018; Samsuddin et al., 2018). We can see a highly temporal correspondence between the mean temperature and the PM_{10} concentration.

From Figure 3, the distribution of the air quality parameters is relatively location-dependent. The yearly mean concentrations of CO, NO_2, and SO_2 have similar regional distribution characteristics. In the middle region of the Peninsula, the concentrations of CO, NO_2, and SO_2 are relatively low and constant throughout the 20-year period. They also show a certain association with the locations of the cities and the population. Especially in the regions of Kuala Lumpur and Bukit Mertajam, the concentrations are high compared to those of the other regions. A small decrease in the CO and NO_2 concentrations is revealed over time. In contrast, the decline in the SO_2 concentration over time is very pronounced, and especially in the early 2000s, the reduction was steep.

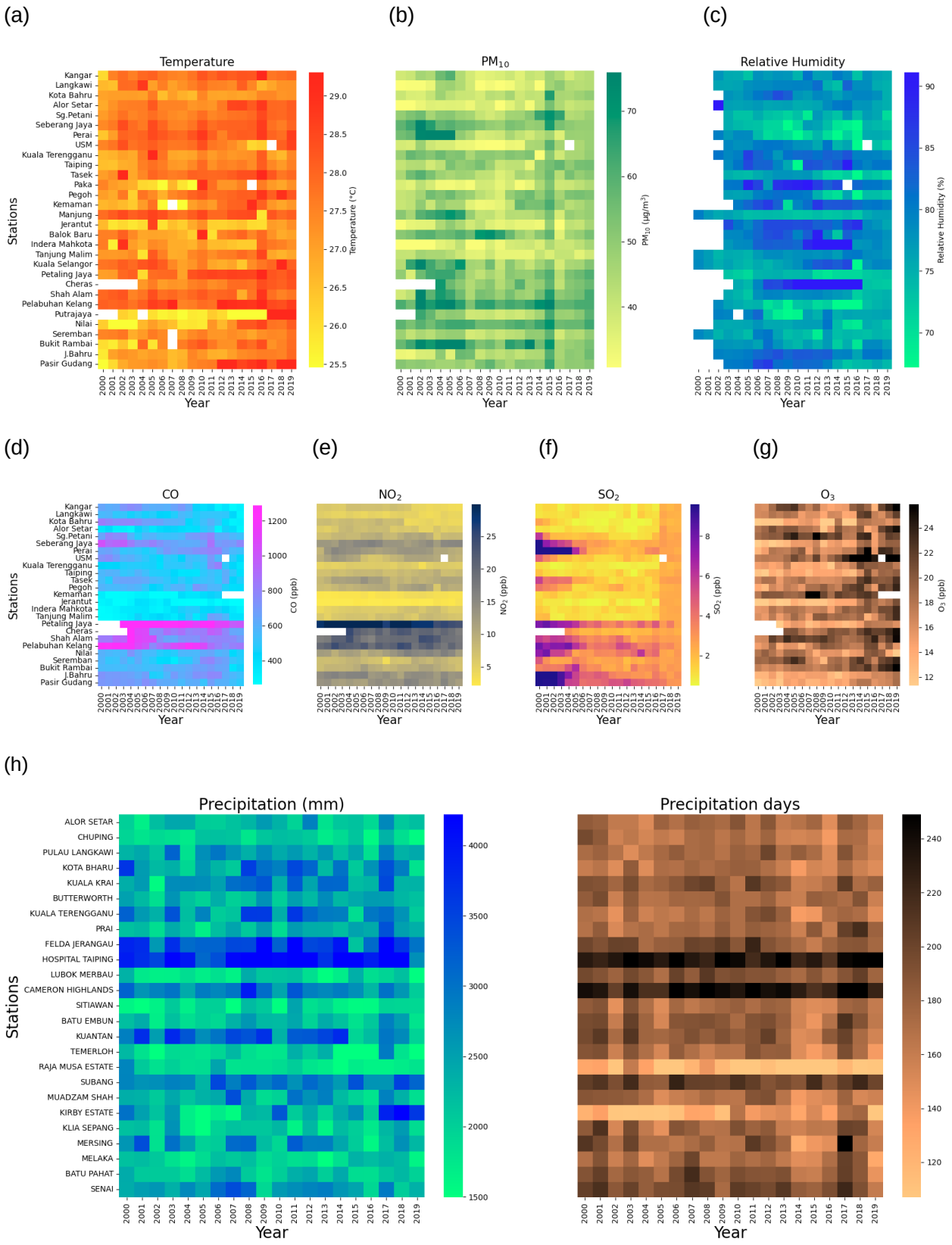


Fig. 3 (a) Yearly mean temperature, (b) PM_{10}, (c) relative humidity, (d) CO, (e) NO_2, (f) SO_2, (g) O_3, and (h) precipitation amount and days with precipitation from north (top) to south (bottom) over the observation period.

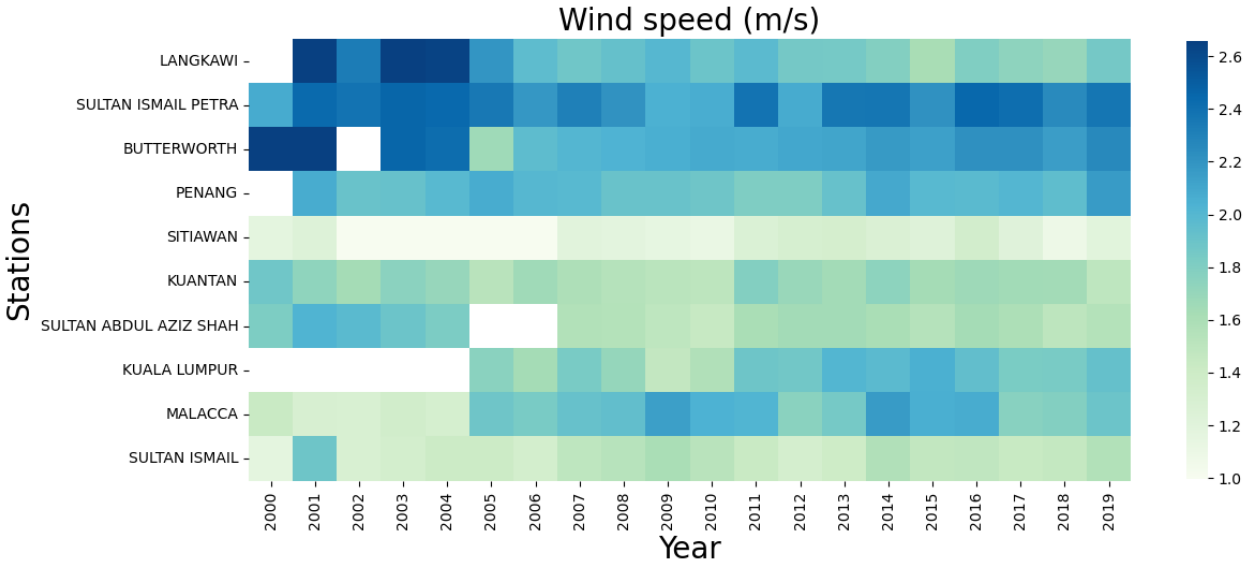
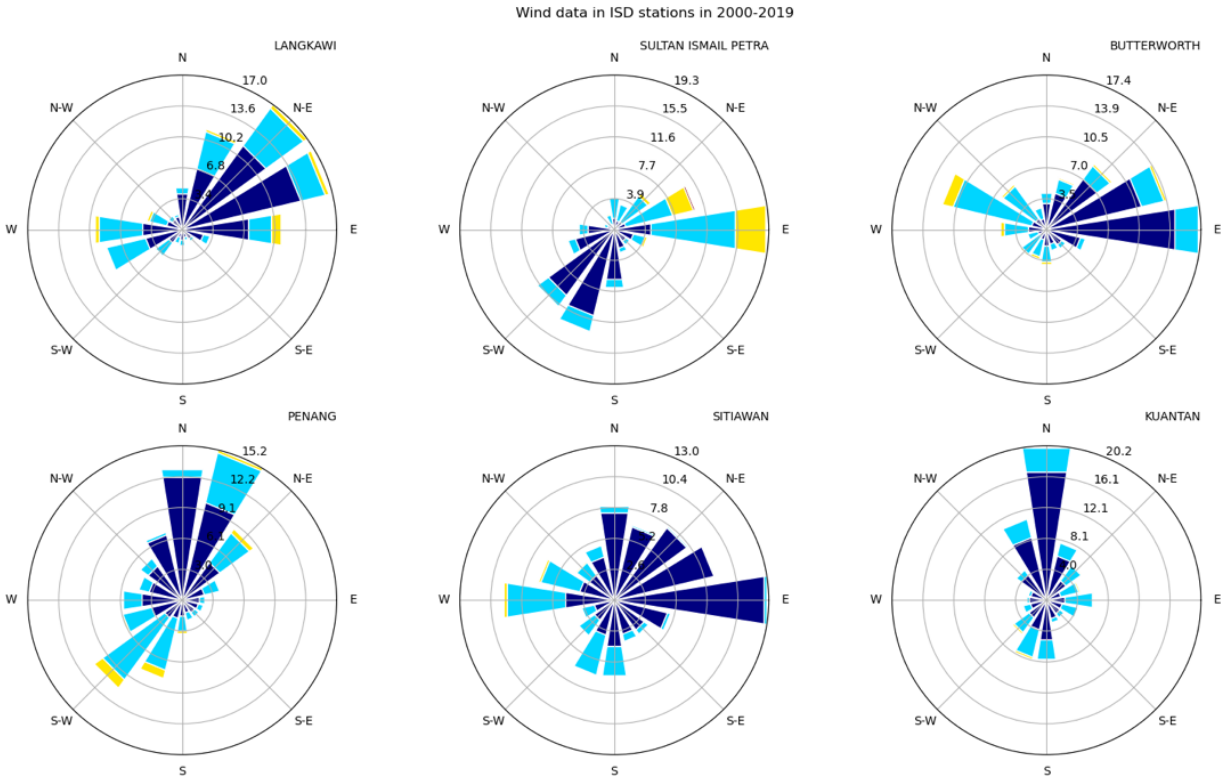


Fig. 4 Yearly mean wind speed from north (top) to south (bottom) over the observation period.



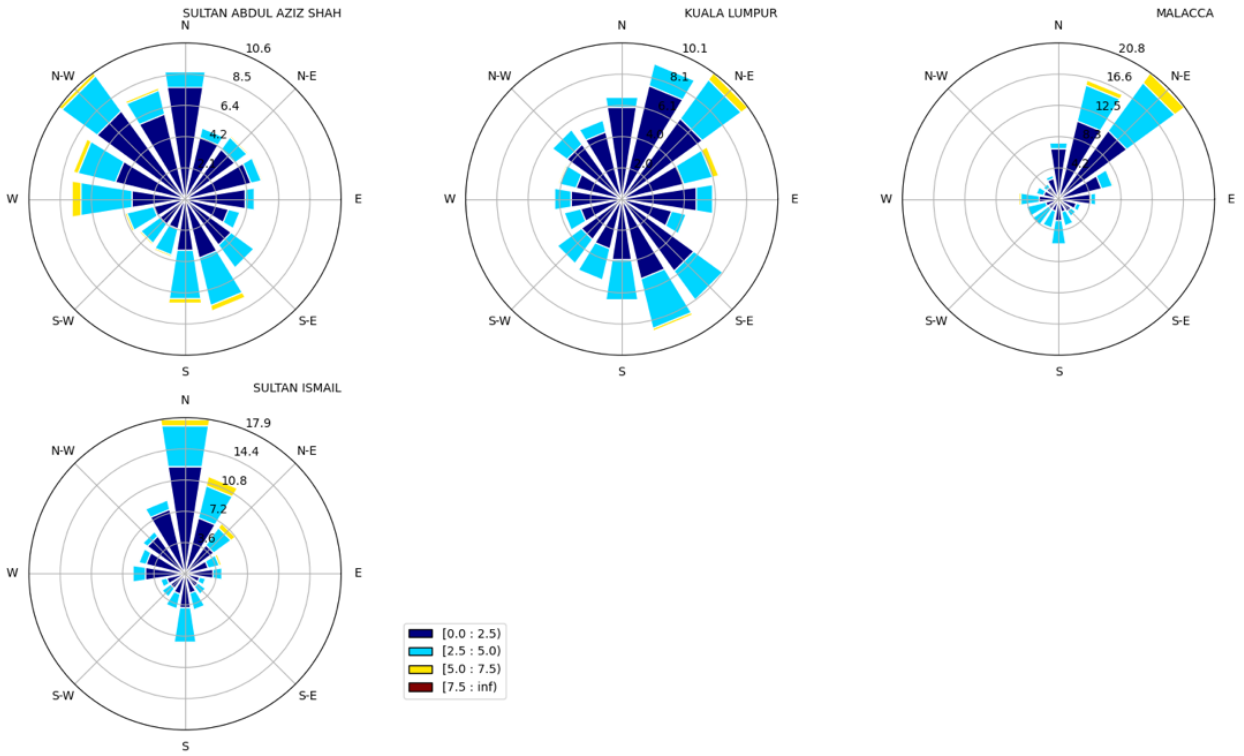


Fig. 5 Wind rose map of the surface hourly wind speed and wind direction from 10 general weather stations over the observation period.

The average concentration of O_3 fluctuates all over the Peninsula. Compared with the other air pollutants studied in this paper, the temporal and spatial variation characteristics of ozone are weaker, but a certain degree of urban agglomeration can still be seen. The ozone concentration generally increased, specifically after 2014, at most of the stations. This leads to the conclusion that the high amount of pollution in the Peninsula varies regionally and temporally. Peninsular Malaysia's central area has the greatest pollution concentration (Sentian et al., 2019).

The geographical distribution of air pollution concentrations obviously highlights urban agglomeration, which also indicates the degree of urbanization. Climate change raises numerous significant air-chemistry problems, and human urban activity is contributing significantly to the current rate of climate change (Mika et al., 2018). Other findings also revealed that economic expansion has a large positive influence on carbon emissions and air pollution (Ali et al., 2017; Dash et al., 2020). Therefore, regional climate change and air quality research is urgent, and paying attention to the balance between the process of urbanization and air quality governance is one of the key points that policymakers need to focus on.

3.3 Monthly variation of climate and air quality

This subsection aims to investigate the broad patterns of monthly variability in climate and air quality conditions across the Peninsular region. Figure 6 shows the monthly variation of variables. Malaysia's climate is divided into four episodes: the inter-monsoon periods in April and October, the southwest monsoon from May to September, and the northeast monsoon from November to March. The action of the monsoon has a great influence on the changes in the monthly patterns. The average monthly temperature reaches the highest value of the year in May (28.2 °C) and then decreases to the lowest value (26.8 °C) within the northeast monsoon period. SO2 has an analogous trend, forming a mountain-like shape; the peak is located around June and July. The precipitation shows an opposite monthly variation compared to O3 and at the same time presents a similar variation compared to the relative humidity. The largest amount of precipitation (332.2 mm) and humidity (66.9%) coincide with the early northeast monsoon season, which is around November. Meanwhile, the lowest amount of precipitation (108.6 mm) and humidity (59.4%) occur in February, when the rainy season ends.

NO_2 and CO both have a tri-peak form. They both peak in April, June/July, and September. The concentration of CO has two peaks that coincide with the two transition periods of the monsoon seasons at very close values, and so does the concentration of NO_2, but with different magnitudes. The high peak is in April (11.3 ppb) and the low peak is in September (10.9 ppb). PM_{10} and O_3 present strong bimodal distributions. Both have their first peak around February and March. The difference is that during the southwest monsoon, the PM_{10} peak reached the maximum PM_{10} value (54.82 $\mu g/m^3$), and O_3 experienced a smaller peak (18 ppb) compared to its first peak (20 ppb).

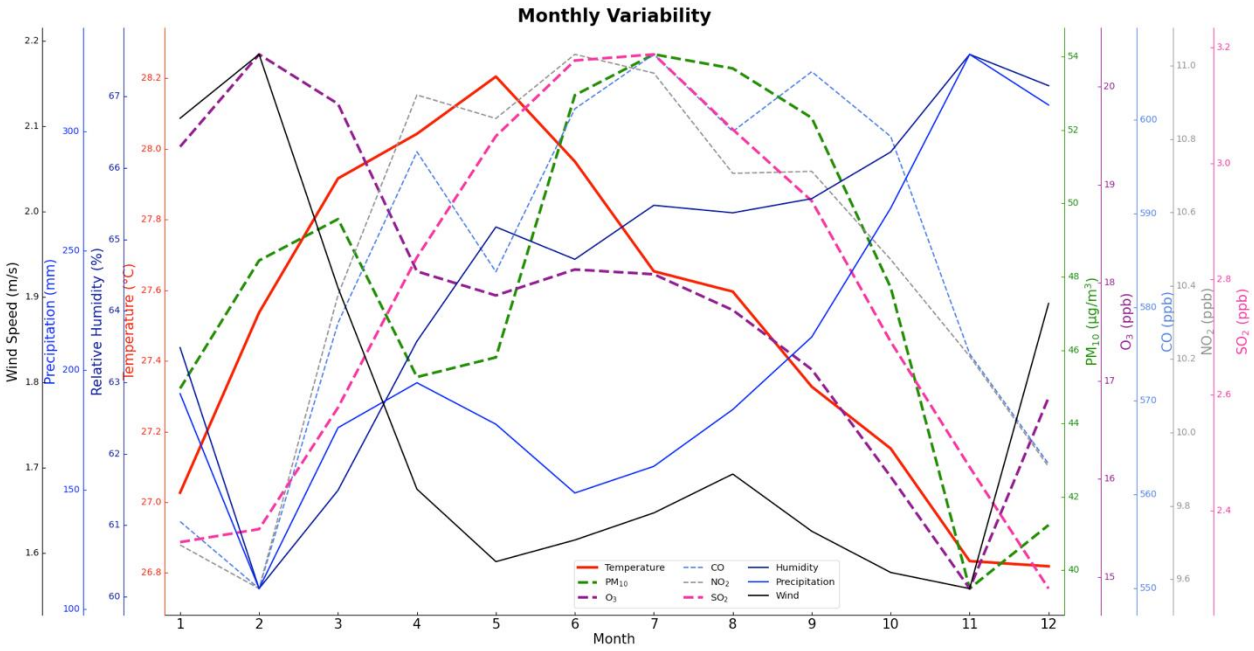


Fig. 6 Monthly variability of mean monthly climate (mean temperature, relative humidity, precipitation, and wind speed; solid lines) and air quality (PM_{10}, O_3, CO, NO_2, and SO_2; dotted lines) over the observation period.

The monthly variation of the wind speed is in good agreement with the monsoon variation. When the northeast monsoon prevails, the wind speed is high, and the monthly average wind speed can reach up to 2.02 m/s. During the southwest monsoon period, the wind speed is relatively low. Winds are often mild and changeable throughout the two inter-monsoon months (Figure 7). Because Malaysia is mostly a maritime country, the impact of land and sea breezes on the overall wind flow pattern is significant, especially on clear days (Hanoon et al., 2022).

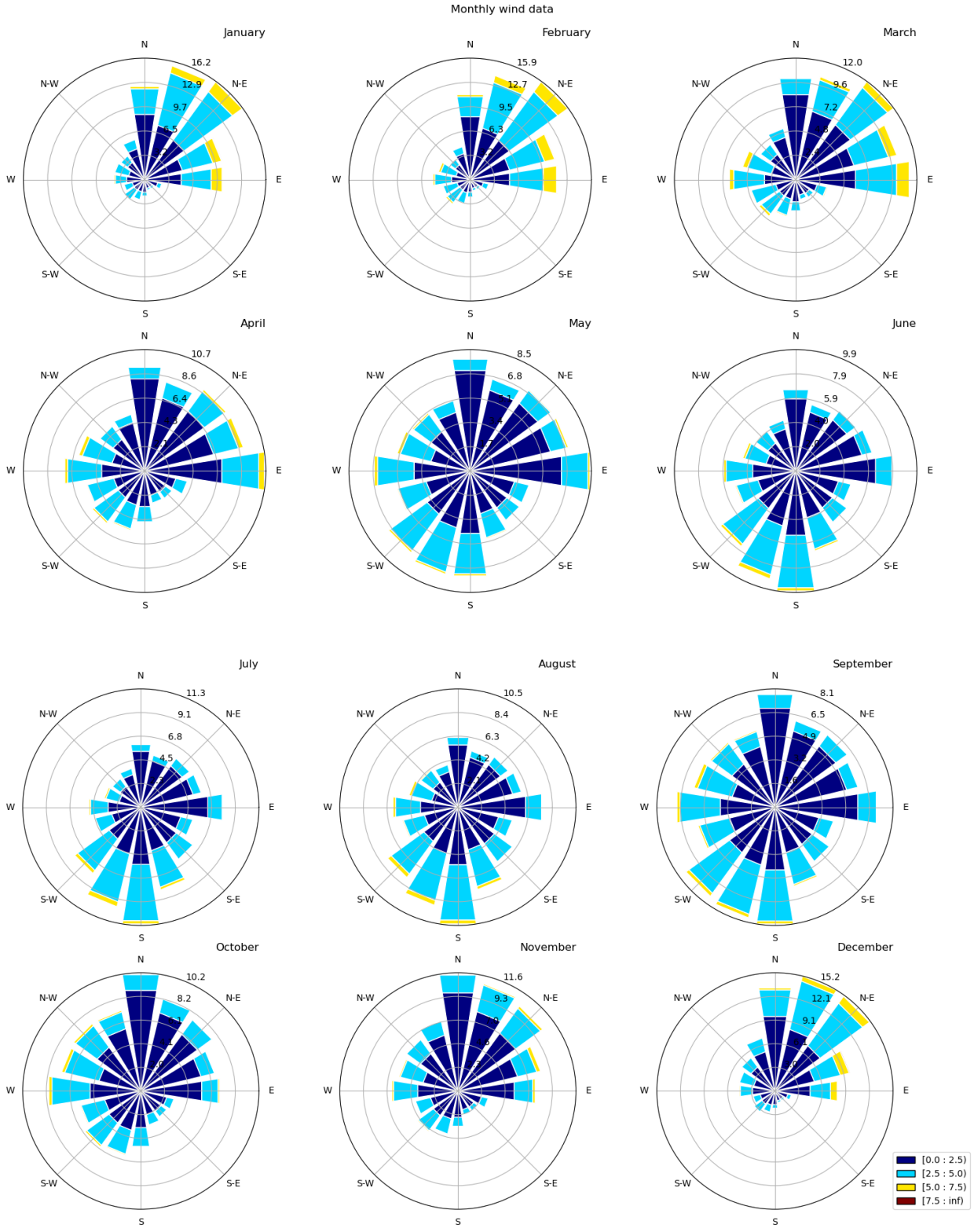


Fig. 7 Seasonality of the monthly wind speed and wind direction over the observation period.

3.4 Diurnal variation of climate and air quality

The diurnal variation of the variables is revealed in Figure 8. Most of the variables show a conspicuous diurnal mode. The hourly average temperature increases at 08:00 local time (LT), peaks around late noon (31.84°C), and then decreases smoothly. The relative humidity has a similar pattern compared with PM_{10}, CO, and NO_2, but without the first trough, creating a horseshoe shape. The wind speed, unlike the other parameters, has a mountain-like pattern, with a peak at 08:00–09:00 LT. Wind speeds are low and steady at night.

PM_{10}, CO, and NO_2 have bimodal modes, with two peaks at around 08:00–10:00 LT and 22:00–23:00 LT and troughs at around 05:00–06:00 LT and 14:00–15:00 LT. The trend of SO_2 is obviously different from those of the other factors. SO_2 reaches its minimum value (2.1 ppb) at around 06:00 LT, and then sharply jumps to the maximum value (3.3 ppb) at 10:00 LT; this is followed by a slight decrease and then an increase in the afternoon and at night. By comparison, the regularity of the daily changes is much higher than that of the monthly and interannual changes.

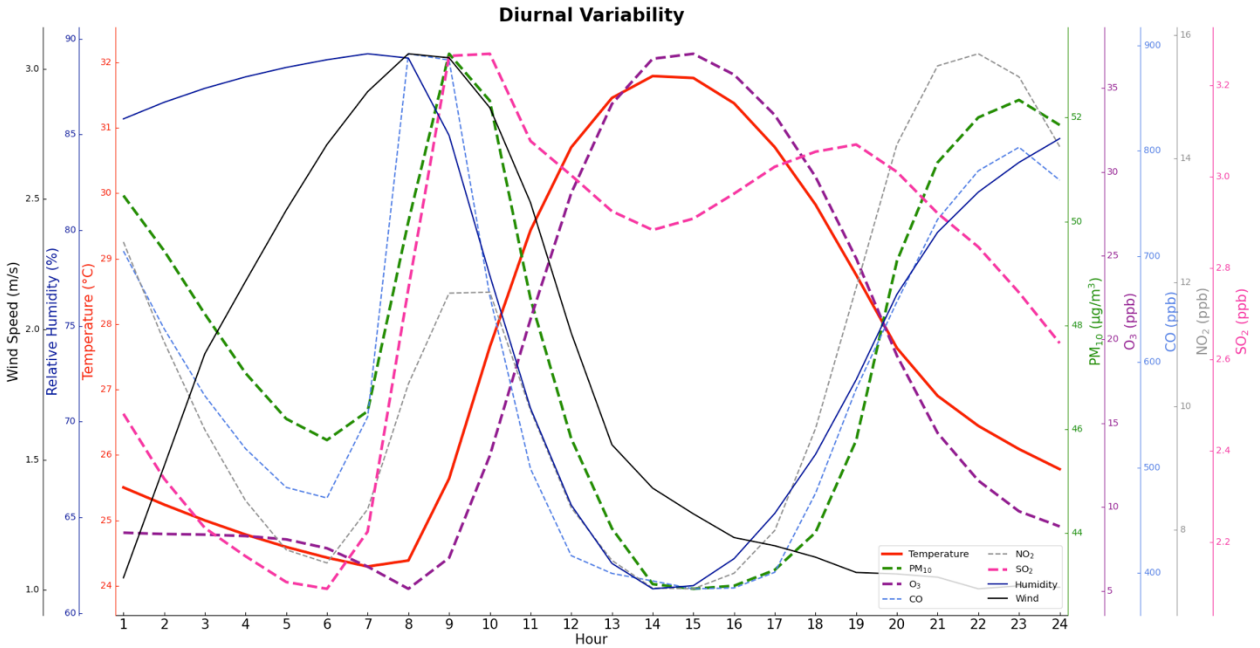
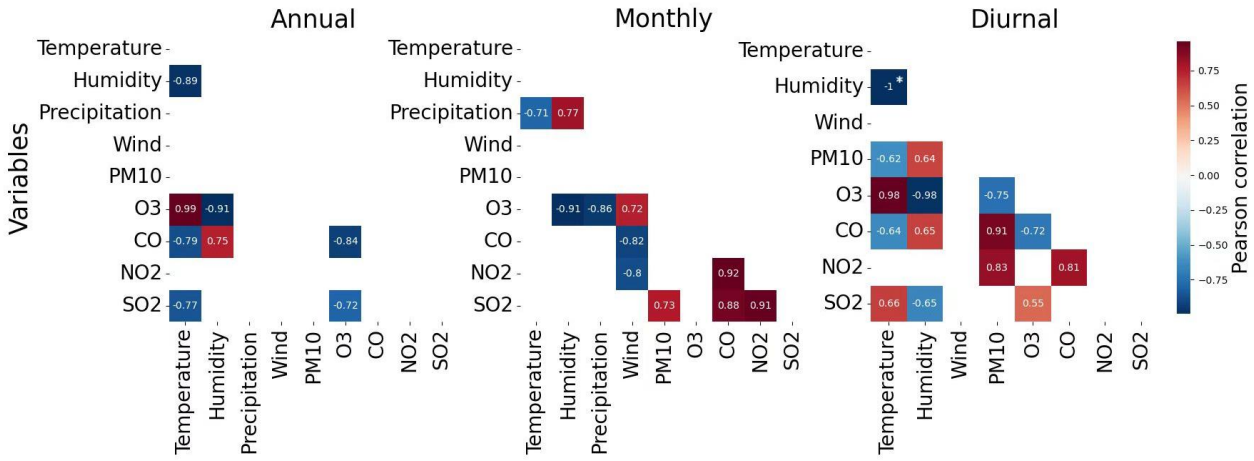


Fig. 8 Diurnal variability of mean hourly climate (mean temperature, relative humidity, precipitation, and wind speed; solid lines) and air quality (PM_{10}, O_3, CO, NO_2, and SO_2; dotted lines) over the observation period.

3.5 Correlation of climate and air quality

In this study, Pearson correlation coefficients were calculated between each climate variable and each air quality variable, respectively, on year-to-year, month-to-month, and day-to-day time scales (Figure 9). The Shapiro-Wilk test for normality was applied to all data (refer to Supplementary S3 for details), and the Pearson correlation coefficients are presented in Figure 9. For interannual variation, the temperature is significantly positively correlated with O_3 (0.99) and strongly negatively correlated with CO (-0.79) and SO_2 (-0.77). On the other hand, SO_2 and CO both have strong negative correlations with O_3. The humidity has strong positive (negative) correlations with CO (O_3). The factors' relationships for monthly changes are more prominent than those for annual changes. On a monthly scale, the temperature is negatively correlated with precipitation. O_3 has a strong negative correlation with the relative humidity (-0.91) and precipitation (-0.86). NO2 is highly correlated with CO (SO2), with a coefficient of 0.92 (0.91). In terms of diurnal variation, prominent correlations among the factors are abundant. The temperature and humidity are negatively correlated. Additionally, the temperature has a strong connection with O3 (0.98), as it does on an annual scale. PM10 is highly correlated with CO (0.91) and NO2 (0.83). O3 is negatively correlated with the relative humidity (-0.98). NO2 and CO also have a high positive correlation. The numbers of significant correlations (0.7 or -0.7) for each time scale are 8 (annual), 11 (monthly), and 8 (daily), respectively. It can be observed that when different time scales are focused on, the correlations undergo significant changes.

The results concerning the linear relationship between climate change factors and air pollutant factors vary widely among previous studies, depending on the size of the region studied, the frequency of the observations, and the time horizon (Ismail et al., 2011; Lim et al., 2022; Miyama et al., 2020; Turalioglu et al., 2005; Suris et al., 2022). A previous study found that for some specific Malaysian stations, the temperature had a positive connection with the PM_{10} concentration but a negative correlation with the relative humidity (Dominick et al., 2012). SO_2 and NO_x levels were negatively correlated with temperature throughout the summer and monsoon seasons but positively correlated with temperature during the pre- and post-monsoon seasons (Jayamurugan et al., 2013). However, in general, there is a positive correlation between the air temperature and ozone content on large time scales. At elevated temperatures, ozone production accelerates and emissions of its natural components increase. The combination of high temperatures and feeble winds causes the atmosphere to stagnate. So, the air simply heats up and ozone levels can accumulate, which is also known as one of the "climate penalty" phenomenon (Chen et al., 2019; Fu & Tian, 2019; Porter & Heald, 2019). The remaining factors need to be compared under the corresponding regional and temporal conditions. The variables in this subsection exhibit a one-to-one correlation with each other, without any interaction or intervention from other parameters. The interactive correlation among the variables will be taken into account using CCA in the next subsection.



*: The true value is -0.999, which is shown as -1 in the image because it has been rounded to two digits.

Fig. 9 Pearson correlation coefficients of the mean temperature, precipitation, relative humidity, wind speed, PM_{10}, O_3, CO, NO_2, and SO_2 over the observation period. (Correlation values are shown only for those pairs of variables that passed the significance test at the 99% confidence level.)

3.6 CCA of meteorological variables and air pollution variables

Based on the Pearson correlation analysis results for the wind speed performance (as depicted in Figure 9), it can be inferred that the impact of the wind speed on the long-term annual average changes in air quality factors is negligible. Consequently, the influence of the wind speed has been excluded from consideration. Additionally, since the impact of precipitation was also found to be weak (refer to Supplementary S4 for details), the variables have been re-analyzed using CCA after the removal of the precipitation factor. The results of this analysis are presented in Tables 3–5 and Figure 10.

The relationships between the annual mean meteorological variables (Set X) and annual air quality variables (Set Y) (Table 3) were tested using CCA. The results (Table 4) showed a strong correlation between the two datasets (CCA: $r = 0.994$, $p < 0.001$), with 71.18% of the variance represented in the model (eigen = 0.988, df = 10).

Table 3

Datasets of climate variables and air quality variables.

Set X (Meteorological)	Temperature	Humidity			
Set Y (Air Quality)	CO	O_3	SO_2	NO_2	PM_{10}

Table 4

Canonical correlation analysis results.

Canonical variables	**Canonical correlation**	**Proportion of variance explained**	**Eigen values**	**Wilks**	**Degree of freedom**	**F**	**P**
Pair 1	0.994	71.18	0.988	0.007	10	28.692	<0.001

Table 5

Canonical loadings of set Y and set X, respectively.

	X_1
Temperature (T2)	-0.998
Humidity (RH)	0.915
	Y_1
O_3	-0.995
CO	0.799
SO_2	0.753
PM_{10}	-0.489
NO_2	0.029

Table 6

Proportion of variance explained by the first pair of canonical variables.

	Set X	Set Y
X_1	91.676	48.694
Y_1	48.134	90.621

From Table 4, it is observed that the first pair of canonical variables was found to be significant after passing the significance test. The correlation coefficient of the first pair of canonical variables is 0.994. The subsequent analysis will be based on the first pair of canonical variables (X_1, Y_1), which are presented in Table 5, using the equations given below:

$$X_1 = -0.998\ T2 + 0.915\ RH, \quad (9)$$

$$Y_1 = -\ 0.489\ PM_{10} - 0.995\ O_3 + 0.799\ CO + 0.029\ NO_2 + 0.753\ SO_2. \quad (10)$$

The proportions of variance explained by X_1 and Y_1 are given in Table 6. The canonical variable X_1 explains 48.694% of the information of the indicators in set Y, and it explains 91.676% of the information of the indicators in set X. The canonical variable Y_1 explains 90.621% of the information of the indicators in set Y and 48.134% of the information of the indicators in set X. Figure 10 summarizes the CCA, with meteorological variables accounting for 71.182% of the information of the air quality factors. The value here is notably smaller than those in Table 4 and Table 6 because all the pairs of canonical variables are considered in the calculation to obtain an overall linear relationship value between the climate change factor group and the air quality factor group. Compared with the results from the CCA with precipitation (Supplementary S4), the interpretation ratio of meteorological factors to air quality factors improved by about 10%. This result shows that for the long-term interannual variability of these factors, changes in temperature and humidity can better predict the degree of future changes in air quality variables.

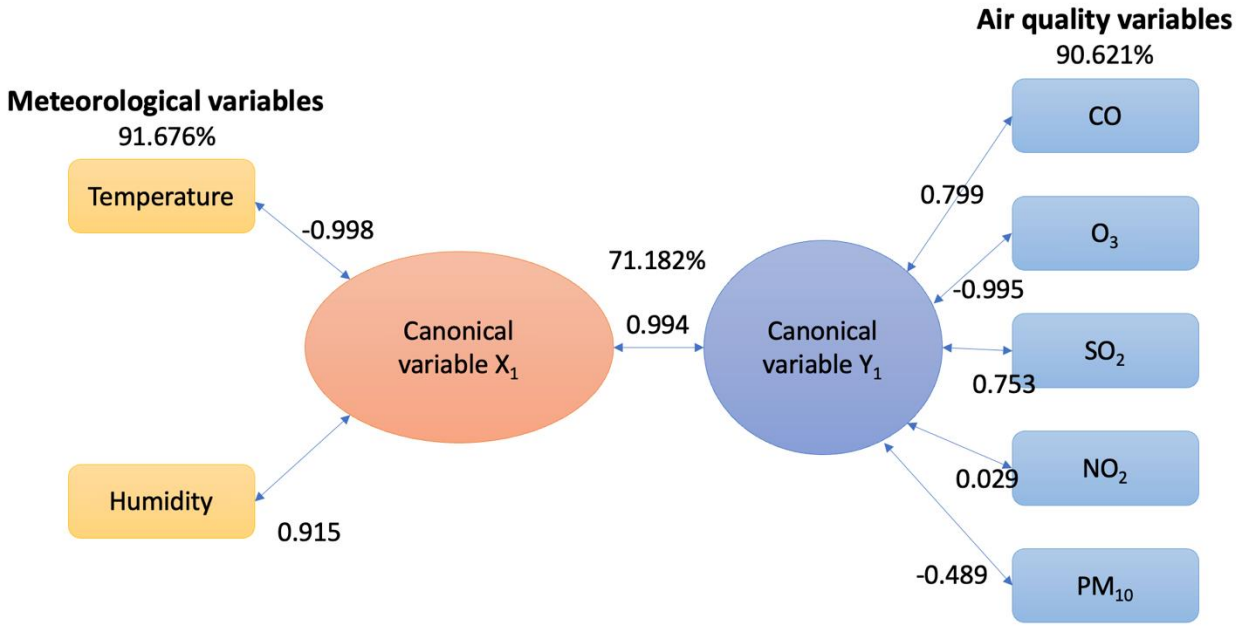


Fig. 10 CCA map of the yearly mean temperature, relative humidity, and PM_{10}, O_3, CO, NO_2, and SO_2 concentrations over the observation period.

3.7 CCA of regional meteorological variables and air quality variables

From the heatmap in subsection 3.1, it is clear that the air quality distribution is relatively location-dependent, with certain regions having a much higher pollution level. Hence, the CCA is conducted on a regional level to determine the level of correlation between the climate and air quality. Taking states as the main basis for distinction, the states of the Peninsula are restructured into eight regions, as shown in Figure 11. The detailed region categorization is shown in Table S5.

An independent CCA study was conducted for each region using the station data within the region. Similar to the previous subsection, only the CCA results without precipitation data are shown here. The results with precipitation data can be found in supplementary information S5. The detailed calculation results are given in Table 7.

Table 7

Regional canonical correlation analysis results, excluding precipitation data.

Region	Canonical variables	Canonical correlation	Proportion of variance explained	Eigenvalue	Wilks	Degree of freedom	F	P
Terengganu	Pair 1	0.937	52.45%	0.878	0.025	10	11.741	<0.001
Johor	Pair 1	0.995	51.27%	0.991	0.001	10	93.416	<0.001
Kedah	Pair 1	0.97	58.67%	0.941	0.02	10	13.416	<0.001
Kelantan	Pair 1	0.97	56.48%	0.941	0.016	10	13.643	<0.001
Pahang	Pair 1	0.931	51.58%	0.867	0.025	10	10.718	<0.001
Perak	Pair 1	0.985	55.78%	0.97	0.007	10	28.736	<0.001
Selangor	Pair 1	0.99	57.24%	0.98	0.005	10	32.581	<0.001
Negeri	Pair 1	0.979	72.82%	0.959	0.026	10	13.42	<0.001
Region	Canonical variables	Canonical correlation	Proportion of variance explained	Eigenvalue	Wilks	Degree of freedom	F	P
Sembilan								

According to Table 7 (Figure 11) and Table S12 (Figure S6), the proportion of variance explained between the climate sets and air quality sets improves by an average of 17.2% after the precipitation data are excluded, especially in the Melaka and Negeri Sembilan regions. Since the urbanization process of the west coast of Malaysia is stronger than that of the east coast, the degree of climate change and air quality pollution are more obvious in densely populated areas (Figure 3). Due to its special geographical location, the west coast area has often been affected by biomass burning from the Indonesian Sumatra and Kalimantan regions, and the temperature, humidity, and air quality are also closely related to the burning situation (Islam et al., 2018; Samsuddin et al., 2018). Thus, based on the graphical comparison, the changes in the air quality

on the west coast are more sensitive and more closely related to climate change than those in the east coast areas, and the predictability is also better. However, the regional canonical correlations of the two cases are not very different.

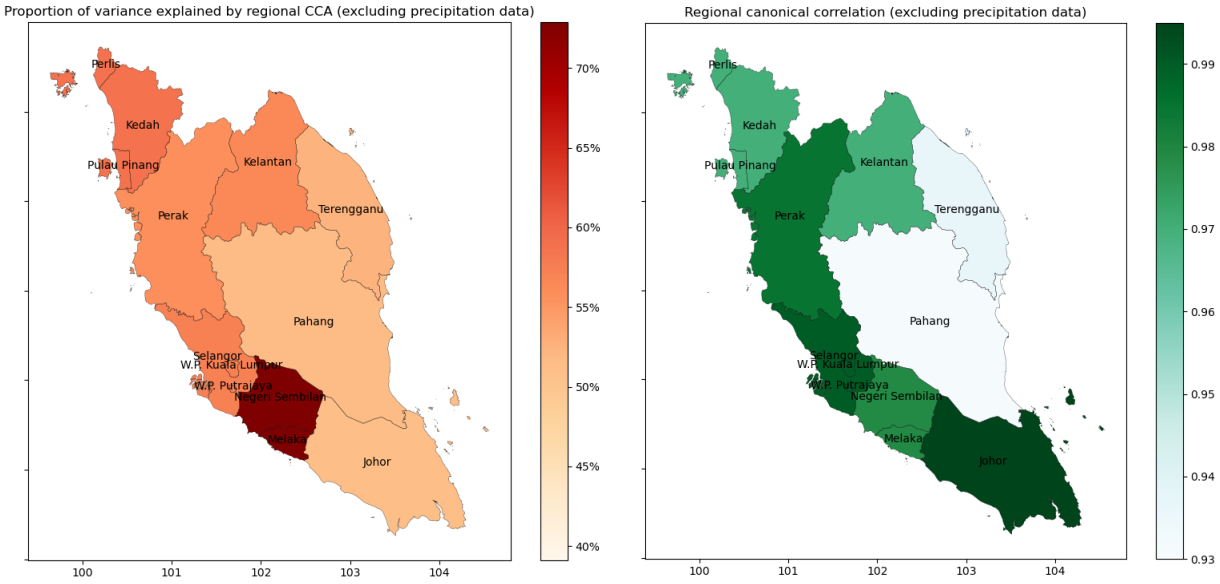


Fig. 11 Proportion of variance explained by regional CCA and regional canonical correlation map of eight regions in the Peninsula over the observation period.

3.7 CCA projection of air quality variables for temperature-rise scenarios

The average worldwide temperature in 2019 was 1.1°C higher than the 1850–1900 average, which is assumed to reflect pre-industrial circumstances (Kappelle, 2020). To better predict the air quality change caused by the temperature change, we used the climate factors without the precipitation data as the dependent variables and carried out the prediction and simulation of the air quality factors using a machine learning-based algorithm from the scikit-learn library in

Python. The temperature is increased by 0.4ºC and 0.9ºC, respectively, as a trigger, and the humidity remains unchanged; 2019 is set as the base year, and the predicted values of the air quality factors are obtained. The results are given in Table 8.

Table 8
Projection results of air quality variables with different increases in temperature.

Temperature increase	PM_{10} ($\mu g/m^3$)	O_3 (ppb)	CO (ppb)	NO_2 (ppb)	SO_2 (ppb)
Value in 2019	51.2	21.3	484	9.6	2.5
Value in 2000	44	15.3	609.6	9.4	5.5
+1.5 ºC	52.2	24.3	452.2	10.4	0.4
Percentage reference based on 2019 (%)	+1.8%	+14.3%	-6.6%	+8%	-86.2%
Percentage reference based on 2000 (%)	+18.5%	+59.4%	-25.8%	+10.3%	-93.6%
Standard deviation	0.4	1.9	2.2	50.1	0.4
+2 ºC	53.7	27.1	393.3	10.4	-0.9
Percentage reference based on 2019 (%)	+4.9%	+27.6%	-18.8%	+7.8%	-135.3%
Percentage reference based on 2000 (%)	+22.1%	+77.9%	-35.5%	+10%	-116.4%
Standard deviation	0.5	2.1	2.6	59.3	0.4

With a temperature increase of 0.4 ºC (based on 2019 data, the relative humidity remains the same), the concentrations of PM_{10}, O_3, and NO_2 all present upward trends. Compared with the 2019 value, ozone increased the most (+14.3%), and CO and SO_2 decreased. When the temperature increased by 0.9 ºC, the values of the above air variables changed more strongly; the increase in the ozone concentration nearly doubled (+27.6%) compared with the previous

scenario, and the reduction in the carbon monoxide concentration more than tripled (-18.8%). Although it is against common sense for the SO_2 concentration to drop to a negative number in the predicted results, we can draw the conclusion that SO_2 will decrease rapidly in the future. The coefficient of determination (R^2 score) of the CCA model for these datasets is 0.47, which indicates that the model is expected to predict future samples with moderate effectiveness. However, we can still get a glimpse of the future from the simulation results; this CCA algorithm's simulation outcomes are also roughly consistent with the findings of earlier studies. The concentration of nitrogen dioxide increases under different carbon emission pathways (Lim et al., 2022), and in the Representative Concentration Pathway RCP4.5 scenario, the concentrations of ozone and PM_{10} also increase (Mika et al., 2018). A slight change in temperature will have a dramatic impact on the concentrations of the air quality factors. In the event of further temperature increases in the future, we should pay more attention to the concentrations of ozone, PM_{10}, and nitrogen dioxide, as these air pollutants are more sensitive to temperature.

In terms of chemistry processes, the positive surface ozone-temperature association is primarily driven by the fact that an increase in temperature boosts natural emissions and increases ozone chemical synthesis at high NOx levels (Gu et al., 2020; Lu et al., 2019). The drop in SO_2 is mostly attributable to the government's energy emission control and clean air policy regulations rather than climate change (Cowern et al., 2018; Ukhov et al., 2020). According to National Center for Atmospheric Research (NCAR) research reports, during a heat wave, the intense heat and stagnant air increase the amount of ozone pollution and particle pollution. Drought

conditions may also exacerbate forest fires and hence contribute to particulate pollution in the atmosphere. NO_2 mostly enters the atmosphere due to fuel combustion and human emissions. NO_2 and other NO_x may react with other airborne molecules to produce both particulate matter and ozone. The long-term growth in CO levels seems to have ceased and reversed over the last several years. This might have occurred for a variety of reasons, such as increased CO removal from the environment and enhanced combustion efficiency in industry. The increased use of natural gas in recent years and regional decreases in human-caused CO emissions have reduced atmospheric CO levels (Campbell et al., 2018; Gratz et al., 2015; WHO, 2021). The NO2 and CO content in the air directly responds to temperature changes, although the chemical process is very weak; changes in the NO2 and CO content are mainly related to the control of emissions and the intervention of human factors.

Since CCA is based on pure linear analysis, it can only reflect the association between variable groups to a certain extent; this has limitations, especially due to the non-linearity of chemical processes. Therefore, to better analyze the correlation between the two sets of variables, and to predict the degree of change in the air quality factors based on changes in the climate factors, numerical model simulation is an indispensable method that must be applied. In subsequent research, it is necessary to carry out regional climate and air quality research using a numerical chemical climate prediction model to resolve the physical and chemical responses of air pollutants to changes in climate conditions.

4 CONCLUSIONS

In this study, temperature, relative humidity, precipitation, wind speed, wind direction, CO, NO_2, O_3, SO_2, and PM_{10} ground-based observational data have been used to assess the temporal and spatial climate and air quality characteristics in the period from 2000–2019 in Peninsular Malaysia. Using Pearson correlation analysis and canonical correlation analysis (CCA), this study identifies the long-term relationships between climate factors and the variability of air quality variables. Using the CCA algorithm based on machine learning, this study forecasts the future air quality under various scenarios of temperature rise. The following summarizes the main findings of this article in bullet form.

- The study reveals a clear warming trend in the climate, with an upward trend in the temperature (+1.14°C) and a slight decrease in the relative humidity, while the precipitation showed significant fluctuations that increased in amplitude.
- The annual average ozone concentration showed a steady increase year by year (+39.5%), while the SO2 concentration showed a steady decline (-53.6%).
- The high correlation between climate variables and air quality variables provides an important basis for predicting future changes in the air quality due to climate change, which makes CCA projection a feasible way to forecast future air quality.
- According to the CCA results, there is a strong correlation between climate and air quality; in the state of Selangor, the sensitivity of the air quality factors to meteorological factors is particularly strong.
- A 2°C increase in temperature caused an increase in PM10 (+22.1%), O3 (+77.9%), and NO2 (+10%), while CO (-35.5%) and SO2 (-116.4%) decreased.

- Slight changes in temperature have a significant impact on air quality, and further temperature increases in the future require greater attention to be paid to the concentrations of O3, PM10 and NO2.
- The long-term data analysis and characterization of climate and air quality will assist policymakers and the relevant authorities in adapting measures and policies to future conditions.

Acknowledgments:

We gratefully acknowledge the Ministry of Higher Education Malaysia for supporting this research through Long-term Research Grant Scheme Project 3, grant number LRGS/1/2020/UKM–UKM/01/6/3, which is under the program of LRGS/1/2020/UKM/01/6. Furthermore, the authors would like to thank the Department of the Environment Malaysia and the Malaysia Meteorological Department for providing the data used in the analysis.

References:

Ahamad, F.G., Paul T., Latif, M.T., Juneng, L., Xiang, C.J., 2020. Ozone trends from two decades of ground level observation in Malaysia. *Atmosphere*. *11*. https://doi.org/10.3390/atmos11070755.

Akbaş, Y., Takma, C., 2005. Canonical correlation analysis for studying the relationship between egg production traits and body weight, egg weight and age at sexual maturity in layers. *Czech J. Anim. Sci.* 50(4), 163–168.

Ali, H.S., Abdul-Rahim, A.S., Ribadu, M.B., 2017. Urbanization and carbon dioxide emissions in Singapore: Evidence from the ARDL approach. *Environ. Sci. Pollut. Res. Int. 24*(2), 1967–1974. https://doi.org/10.1007/s11356-016-7935-z.

Andrée, B. P. J., Chamorro, A., Spencer, P., Koomen, E., & Dogo, H. (2019). Revisiting the relation between economic growth and the environment; a global assessment of deforestation, pollution and carbon emission. *Renewable and Sustainable Energy Reviews*, *114*, 109221.

Bowo, A. M., Irianingsih, I., & Ruchjana, B. N. (2020). Canonical Correlation Analysis of Global Climate Elements and Rainfall in the West Java Regions. *Desimal: Jurnal Matematika*, *3*(2), 143-154. https://doi.org/10.24042/djm.v3i2.5870.

Campbell, P., Zhang, Y., Yan, F., Lu, Z., Streets, D., 2018. Impacts of transportation sector emissions on future U.S. air quality in a changing climate. Part I: Projected emissions, simulation design, and model evaluation. *Environ. Pollut*. *238*, 903–917. https://doi.org/10.1016/j.envpol.2018.04.020

Chen, Z., Zhuang, Y., Xie, X., Chen, D., Cheng, N., Yang, L., & Li, R. (2019). Understanding long-term variations of meteorological influences on ground ozone concentrations in Beijing During 2006–2016. *Environmental pollution*, *245*, 29-37.

Cowern, N.E.B., 2018. Current rapid global temperature rise linked to falling SO2 emissions. *Earth Syst. Dynam. Discuss.* https://doi.org/10.5194/esd-2018-83.

Dash, D.P., Behera, S.R., Rao, D.T., Sethi, N., Loganathan, N., Ercolano, S., 2020. Governance, urbanization, and pollution: A cross-country analysis of global south region. *Cogent Economics & Finance*. *8*(1). https://doi.org/10.1080/23322039.2020.1742023.

Dey Sagnik, C.S., 2018. Air quality in changing climate: Implications for health impacts, in: *Climate Change and Air Pollution.* pp. 9–24. https://doi.org/10.1007/978-3-319-61346-8_2.

Dominick, D., Juahir, H., Latif, M.T., Zain, S.M., Aris, A.Z., 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric Environment*. *60*, 172–181. https://doi.org/10.1016/j.atmosenv.2012.06.021.

Fu, T.-M., Tian, H., 2019. Climate change penalty to ozone air quality: Review of current understandings and knowledge gaps. *Current Pollution Reports*. *5*(3), 159–171. https://doi.org/10.1007/s40726-019-00115-6.

Gratz, L.E., Jaffe, D.A., Hee, J.R., 2015. Causes of increasing ozone and decreasing carbon monoxide in springtime at the Mt. Bachelor Observatory from 2004 to 2013. *Atmospheric Environment*. *109*, 323–330. https://doi.org/10.1016/j.atmosenv.2014.05.076.

Gu, Y., Li, K., Xu, J., Liao, H., Zhou, G., 2020. Observed dependence of surface ozone on increasing temperature in Shanghai, China. *Atmospheric Environment*. *221*. https://doi.org/10.1016/j.atmosenv.2019.117108.

Halim, N.D.A., Latif, M.T., Ahamad, F., Dominick, D., Chung, J.X., Juneng, L., Khan, M.F., 2018. The long-term assessment of air quality on an island in Malaysia. *Heliyon*. *4*(12), e01054. https://doi.org/10.1016/j.heliyon.2018.e01054.

Hanif, M.F., Mustafa, M.R.U., Liaqat, M.U., Hashim, A.M., Yusof, K.W., 2022. Evaluation of long-term trends of rainfall in Perak, Malaysia. *Climate*. *10*(3). https://doi.org/10.3390/cli10030044.

Hanoon, M.S., Ahmed, A.N., Kumar, P., Razzaq, A., Zaini, N.a., Huang, Y.F., Sherif, M., Sefelnasr, A., Chau, K.w., El-Shafie, A., 2022. Wind speed prediction over Malaysia using various machine learning models: Potential renewable energy source. *Engineering Applications of Computational Fluid Mechanics*. *16*(1), 1673–1689. https://doi.org/10.1080/19942060.2022.2103588.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J.J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., Suttie, M.,

2019. The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*. *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019.

Intergovernmental Panel on Climate Change (IPCC). (2014). Near-term Climate Change: Projections and Predictability. In *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 953-1028). Cambridge: Cambridge University Press. doi:10.1017/CBO9781107415324.023

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Islam, M., Chan, A., Ashfold, M., Ooi, C., Azari, M., 2018. Effects of El-Niño, Indian Ocean Dipole, and Madden-Julian Oscillation on surface air temperature and rainfall anomalies over Southeast Asia in 2015. *Atmosphere*. *9*(9). https://doi.org/10.3390/atmos9090352.

Ismail, M., Suroto, A., Ismail, N.A., 2011. Time series analysis of surface ozone monitoring records in Kemaman, Malaysia, in: Haryanto, B. (Ed.), *Air Pollution – A Comprehensive Perspective*. IntechOpen.

Jayamurugan, R., Kumaravel, B., Palanivelraja, S., Chockalingam, M.P., 2013. Influence of temperature, relative humidity and seasonal variability on ambient air quality in a coastal urban area. *International Journal of Atmospheric Sciences*. *2013*, 1–7. https://doi.org/10.1155/2013/264046.

Jud, M.A., Juneng, L., Tangang, F.T., Latif, M.T., Chung, J.X., Ahamad, F., 2020. Madden Julian Oscillation modulation for surface ozone in Peninsular Malaysia. *Atmospheric Environment*. *233*, 117577. https://doi.org/10.1016/j.atmosenv.2020.117577.

Juneng, L., Latif, M. T., Tangang, F.T., Mansor, H., 2009. Spatio-temporal characteristics of PM10 concentration across Malaysia. *Atmospheric Environment*. *43*(30), 4584–4594. https://doi.org/10.1016/j.atmosenv.2009.06.018.

Kappelle, Maarten. (2020). WMO Statement on the State of the Global Climate in 2019. 10.13140/RG.2.2.13705.19046.

Khoir, A.N.u., Ooi, M.C.G., Juneng, L., Ramadhan, M.A.I., Virgianto, R.H., Tangang, F., 2022. Spatio-temporal analysis of aerosol optical depth using rotated empirical orthogonal function over the Maritime Continent from 2001 to 2020. *Atmospheric Environment*. *290*, 119356. https://doi.org/10.1016/j.atmosenv.2022.119356.

Kuwata, M., Miyakawa, T., Yokoi, S., Khan, M.F., Latif, M.T., 2021. The Madden‐Julian Oscillation modulates the air quality in the Maritime Continent. *Earth and Space Science*. *8*(7). https://doi.org/10.1029/2021ea001708.

Kwan, M.S., Tangang, F.T., Juneng, L. 2013. Projected changes of future climate extremes in Malaysia. *Sains Malaysiana*. *42*(8), 1051–1059.

Langworthy, B.W., Stephens, R.L., Gilmore, J.H., Fine, J.P., 2021. Canonical correlation analysis for elliptical copulas. *Journal of Multivariate Analysis*. *183*, 104715. https://doi.org/10.1016/j.jmva.2020.104715.

Lim, N.O., Hwang, J., Lee, S.J., Yoo, Y., Choi, Y., Jeon, S., 2022. Spatialization and prediction of seasonal NO2 pollution due to climate change in the Korean capital area through land use regression modeling. *Int. J. Environ. Res. Public Health*. *19*(9). https://doi.org/10.3390/ijerph19095111.

Ling, M.K.S.a.L.K., 2009. Regional climate observation and simulation of extreme temperature and precipitation trends.

Lu, X., Zhang, L., Shen, L., 2019. Meteorology and climate influences on tropospheric ozone: A review of natural sources, chemistry, and transport patterns. *Current Pollution Reports*. *5*(4), 238–260. https://doi.org/10.1007/s40726-019-00118-3.

Malaysia, J.M., 2009. *Climate Change Scenarios for Malaysia, 2001-2099: Malaysian Meteorological Department Scientific Report*. Malaysian Meteorological Department. https://books.google.com/books?id=hdI1nQEACAAJ.

Marzuki, M., Suryanti, K., Yusnaini, H., Tangang, F., Muharsyah, R., Vonnisa, M., Devianto, D., 2021. Diurnal variation of precipitation from the perspectives of precipitation amount, intensity and duration over Sumatra from rain gauge observations. *International Journal of Climatology*. *41*(8), 4386–4397. https://doi.org/10.1002/joc.7078.

Mayowa, O.O., Pour, S.H., Shahid, S., Mohsenipour, M., Harun, S.B., Heryansyah, A., Ismail, T., 2015. Trends in rainfall and rainfall-related extremes in the east coast of Peninsular Malaysia. *Journal of Earth System Science*. *124*, 1609–1622. https://doi.org/10.1007/s12040-015-0639-9.

Mika, J., Forgo, P., Lakatos, L., Olah, A.B., Rapi, S., Utasi, Z., 2018. Impact of 1.5 K global warming on urban air pollution and heat island with outlook on human health effects. *Current Opinion in Environmental Sustainability*. *30*, 151–159. https://doi.org/10.1016/j.cosust.2018.05.013.

Mikhaylov, A., Moiseev, N., Aleshin, K., & Burkhardt, T. (2020). Global climate change and greenhouse effect. *Entrepreneurship and Sustainability Issues*, *7*(4), 2897.

Miyama, T., Matsui, H., Azuma, K., Minejima, C., Itano, Y., Takenaka, N., Ohyama, M., 2020. Time series analysis of climate and air pollution factors associated with atmospheric

nitrogen dioxide concentration in Japan. *Int. J. Environ. Res. Public Health*. *17*(24). https://doi.org/10.3390/ijerph17249507.

Mohtar, A.A.A., Latif, M.T., Baharudin, N.H., Ahamad, F., Chung, J.X., Othman, M., Juneng, L., 2018. Variation of major air pollutants in different seasonal conditions in an urban environment in Malaysia. *Geoscience Letters*. *5*(1). https://doi.org/10.1186/s40562-018-0122-y.

Morrissey, K., Chung, I., Morse, A., Parthasarath, S., Roebuck, M., Tan, M., Wood, A., Wong, P.-F., Frostick, S., 2021. The effects of air quality on hospital admissions for chronic respiratory diseases in Petaling Jaya, Malaysia, 2013–2015. *Atmosphere*. *12*(8). https://doi.org/10.3390/atmos12081060.

Neagu, O., & Teodoru, M. C. (2019). The relationship between economic complexity, energy consumption structure and greenhouse gas emission: Heterogeneous panel evidence from the EU countries. *Sustainability*, *11*(2), 497.

Ng, C.Y., Wan Jaafar, W.Z., Mei, Y., Othman, F., Lai, S.H., Liew, J., 2022. Assessing the changes of precipitation extremes in Peninsular Malaysia. *International Journal of Climatology*. *42*(15), 7914–7937. https://doi.org/10.1002/joc.7684.

Rana, S., Renwick, J., McGregor, J., & Singh, A. (2018). Seasonal Prediction of Winter Precipitation Anomalies over Central Southwest Asia: A Canonical Correlation Analysis Approach. *Journal of Climate*, *31*(2), 727-741. https://doi.org/10.1175/jcli-d-17-0131.1.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. *12*, 2825–2830.

Porter, W. C., & Heald, C. L. (2019). The mechanisms and meteorological drivers of the summertime ozone–temperature relationship. *Atmospheric Chemistry and Physics, 19*(21), 13367-13381.

Samsuddin, N.A.C., Khan, M.F., Maulud, K.N.A., Hamid, A.H., Munna, F.T., Rahim, M.A.A., Latif, M.T., Akhtaruzzaman, M., 2018. Local and transboundary factors' impacts on trace gases and aerosol during haze episode in 2015 El Nino in Malaysia. *Sci. Total Environ*. *630*, 1502–1514. https://doi.org/10.1016/j.scitotenv.2018.02.289.

Schuur, E. A., Abbott, B. W., Bowden, W. B., Brovkin, V., Camill, P., Canadell, J. G., ... & Zimov, S. A. (2013). Expert assessment of vulnerability of permafrost carbon to climate change. *Climatic Change*, 119, 359-374.

Seinfeld, J.H., Pandis, S.N., 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.

Sentian, J., Herman, F., Yih, C.Y., Hian Wui, J.C., 2019. Long-term air pollution trend analysis in Malaysia. *International Journal of Environmental Impacts: Management, Mitigation and Recovery*. *2*(4), 309–324. https://doi.org/10.2495/ei-v2-n4-309-324.

Turalioglu, F.S., Nuhoglu, A., Bayraktar, H., 2005. Impacts of some meteorological parameters on SO2 and TSP concentrations in Erzurum, Turkey. *Chemosphere*. *59*(11), 1633–1642. https://doi.org/10.1016/j.chemosphere.2005.02.003.

Suhaila, J., Yusop, Z., 2018. Trend analysis and change point detection of annual and seasonal temperature series in Peninsular Malaysia. *Meteorology and Atmospheric Physics*. *130*(5), 565–581. https://doi.org/10.1007/s00703-017-0537-6.

Suris, F.N.A., Bakar, M.A.A., Ariff, N.M., Mohd Nadzir, M.S., Ibrahim, K., 2022. Malaysia PM10 air quality time series clustering based on dynamic time warping. *Atmosphere*. *13*(4). https://doi.org/10.3390/atmos13040503.

Svensson, C., Jakob, D.r., Reed, D.W., 2002. Diurnal characteristics of heavy precipitation according to weather type at an upland site in Scotland. *International Journal of Climatology*. *22*(5), 569–585. https://doi.org/10.1002/joc.672.

Tang, K.H.D., 2019. Climate change in Malaysia: Trends, contributors, impacts, mitigation and adaptations. *Sci. Total Environ.* 650 (Pt 2), 1858–1871. https://doi.org/10.1016/j.scitotenv.2018.09.316.

Tangang, F.T., Juneng, L., Ahmad, S., 2006. Trend and interannual variability of temperature in Malaysia: 1961–2002. *Theoretical and Applied Climatology*. *89*(3-4), 127–141. https://doi.org/10.1007/s00704-006-0263-3.

Tangang, F., Farzanmanesh, R., Mirzaei, A., Salimun, E., Jamaluddin, A.F., Juneng, L., 2017. Characteristics of precipitation extremes in Malaysia associated with El Niño and La Niña events. *International Journal of Climatology*. *37*, 696–716. https://doi.org/10.1002/joc.5032.

Ukhov, A., Mostamandi, S., Krotkov, N., Flemming, J., Silva, A., Li, C., Fioletov, V., McLinden, C., Anisimov, A., Alshehri, Y.M., Stenchikov, G., 2020. Study of SO2 pollution in the Middle East using MERRA‐2, CAMS data assimilation products, and high‐resolution WRF‐Chem simulations. *Journal of Geophysical Research: Atmospheres*. *125*(6). https://doi.org/10.1029/2019jd03199.3

WHO, 2021. WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.

Wong, C.-L., 2018. Trend of daily rainfall and temperature in Peninsular Malaysia based on gridded data set. *International Journal of GEOMATE*. *14*(44). https://doi.org/10.21660/2018.44.3707.

Wang, H.-T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2020). Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *NeuroImage*, *216*, 116745.

Xiao, H.M., Lo, M.H., Yu, J.Y., 2022. The increased frequency of combined El Nino and positive IOD events since 1965s and its impacts on maritime continent hydroclimates. *Sci. Rep*. *12*(1), 7532. https://doi.org/10.1038/s41598-022-11663-1.

Yang, X., Liu, W., Liu, W., & Tao, D. (2019). A survey on canonical correlation analysis. *IEEE Transactions on Knowledge and Data Engineering*, *33*(6), 2349-2368.

Yatim, A.N.M., Latif, M.T., Ahamad, F., Khan, M.F., Nadzir, M.S.M., Juneng, L., 2019. Observed trends in extreme temperature over the Klang Valley, Malaysia. *Advances in Atmospheric Sciences*. *36*(12), 1355–1370. https://doi.org/10.1007/s00376-019-9075-0.

Zhang, J., Tourian, M.J., Sneeuw, N., 2020. Identification of ENSO signature in the boreal hydrological cycle through canonical correlation with sea surface temperature anomalies. *International Journal of Climatology*. *40*(15), 6219–6241. https://doi.org/10.1002/joc.6573.

Zhuang, X., Yang, Z., & Cordes, D. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping*, *41*(13), 3807-3833.

Author Contributions

Yijing Zheng: Methodology, Software, Data curation, Writing – Original draft. **Maggie Chel Gee Ooi**: Conceptualization, Writing – Review & editing, Supervision, **Juneng Liew**: Methodology, Supervision, Funding acquisition. **Hin Boo Wee:** Methodology, Writing – Review & editing. **Mohd Talib Latif**: Writing – Review & editing. **Mohd Shahrul Mohd Nadzir**: Writing – Review & editing, **Norfazrin Mohd Hanif**: Data acquisition, Writing – Review & editing **.Andy Chan**: Writing – Review & editing. **Li Li**: Writing – Review & editing. **Norfazilah Ahmad**: Funding acquisition. **Fredolin Tangang**: Funding acquisition, Writing – Review & editing.

Supplementary Information:

S1: Calibration process

The calculation formulas for correction are shown in equation 1-3. In the formula, Mi, Oi represent simulated value of record i, observed value of record i.

Mean bias (MB) = $\frac{1}{N}\ \sum_{i=1}^{N}(Oi - Mi)$ Equation 1

Mean bias between 2 data sets = MB(ASMA data) − MB(PSTW data) Equation 2

Calibration PSTW data = PSTW original data + Mean bias between 2 data sets Equation 3

Most of the chemical species in CAMS Global data are archived as mass mixing ratios (MMR, kg of gas / kg of air). To compare the data with observational data, unit conversion is necessary. The equations that used to convert MMR to volume mixing ratio (VMR, units ppmv or ppbv) for different chemical species were shown in equation 4-7.

To convert data from MMR to VMR you only need molar masses of dry air and molar mass of the atmospheric species:

$VMR(O_3) = 28.9644 / 47.9982 * 1e6 * MMR(O_3)$ Equation 4

$VMR(CO) = 28.9644 / 28.0101 * 1e6 * MMR(CO)$ Equation 5

$VMR(NO_2) = 28.9644 / 46.0055 * 1e6 * MMR(NO_2)$ Equation 6

$VMR(SO_2) = 28.9644 / 64.0638 * 1e6 * MMR(SO_2)$ Equation 7

The 1e9 in the formulae gives parts per million (ppmv).

Since the CAMS global reanalysis (EAC4) data are only available after 2003, so the data analysis in the following are all based on the time series of 2003-2019. The tables below, show mean bias of yearly mean concentration and calculated calibration results of different air quality factors respectively. In the following figures, two calibration methods were compared. The first one is to calculate the mean bias between ASMA 2003-2017 data and the mean bias between PSTW 2017-2019 data, then calculate the difference between the two. The second method is to compute the difference of mean bias between ASMA 2017 data and PSTW 2017 data. The data before and after those two alteration methods are presented in PM_{10}, O_3 and SO_2 air quality variables as representatives in this research.

Table S1. Mean bias of yearly mean SO_2 concentration and calibration results.

SO_2	ASMA 2003	ASMA 2004	ASMA 2005	ASMA 2006	ASMA 2007	ASMA 2008	ASMA 2009	ASMA 2010	ASMA 2011
MB	0.002758	0.002258	0.001649	0.001166	0.000876	0.000913	0.000721	0.000706	0.000871
SO_2	ASMA 2012	ASMA 2013	ASMA 2014	ASMA 2015	ASMA 2016	ASMA 2017	PSTW 2017	PSTW 2018	PSTW 2019
MB	0.000653	0.000509	0.000747	0.000605	0.000861	0.001312	-0.000142	-0.000151	-0.000163
	MB(ASMA 2003-2017 mean)-MB(PSTW 2017-2019 mean)					MB(ASMA 2017)-MB(PSTW 2017)			
Calibration	0.0012588785095					0.0014537407292			

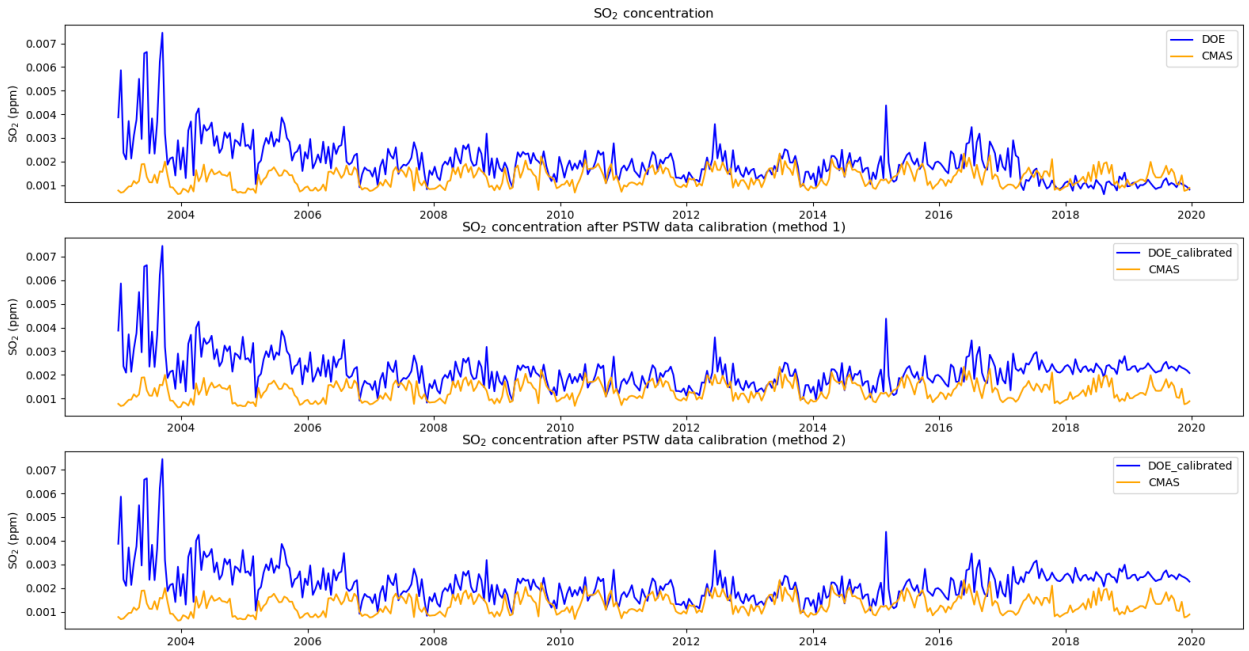


Fig. S1. a) Time series graph of SO_2 concentration data before calibration in 2003-2019. b) Time series graph of SO_2 concentration data after calibration for the first method. c) Time series graph of SO_2 concentration data after calibration for the second method.

Table S2. Mean bias of yearly mean O_3 concentration and calibration results.

O_3	ASMA 2003	ASMA 2004	ASMA 2005	ASMA 2006	ASMA 2007	ASMA 2008	ASMA 2009	ASMA 2010	ASMA 2011
MB	-0.011430	-0.009926	-0.009527	-0.009845	-0.007938	-0.007467	-0.009193	-0.007343	-0.008435
O_3	ASMA 2012	ASMA 2013	ASMA 2014	ASMA 2015	ASMA 2016	ASMA 2017	PSTW 2017	PSTW 2018	PSTW 2019
MB	-0.007763	-0.007183	-0.006251	-0.007649	-0.006021	-0.007395	-0.009341	-0.008232	-0.007345
	MB(ASMA 2003-2017 mean)-MB(PSTW 2017-2019 mean)					MB(ASMA 2017)-MB(PSTW 2017)			
Calibration	0.0000814421770					0.0019459952646			

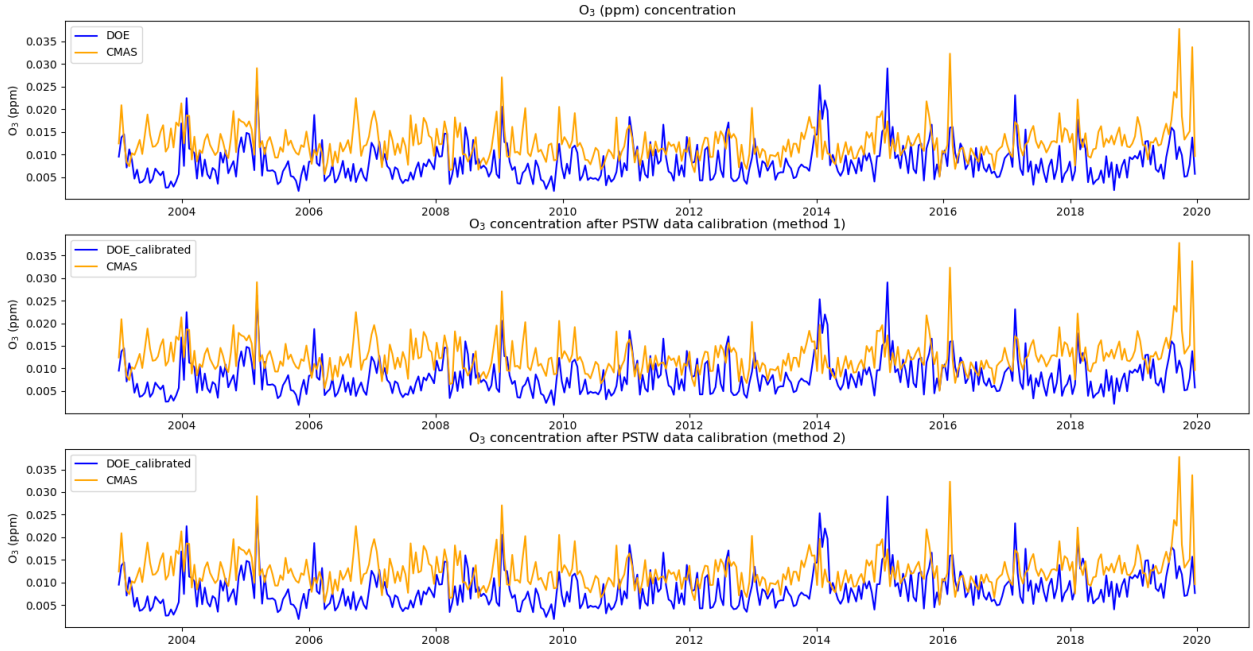


Fig. S2. a) Time series graph of O_3 concentration data before calibration in 2003-2019. b) Time series graph of O_3 concentration data after calibration for the first method. c) Time series graph of O_3 concentration data after calibration for the second method.

Table S3. Mean bias of yearly mean PM_{10} concentration and calibration results.

PM_{10}	ASMA 2003	ASMA 2004	ASMA 2005	ASMA 2006	ASMA 2007	ASMA 2008	ASMA 2009	ASMA 2010	ASMA 2011
MB	10.965302	11.063567	10.988152	9.665828	8.203222	8.381892	7.748867	6.384681	8.895474
PM_{10}	ASMA 2012	ASMA 2013	ASMA 2014	ASMA 2015	ASMA 2016	ASMA 2017	PSTW 2017	PSTW 2018	PSTW 2019
MB	7.967413	6.876545	3.492488	11.726614	9.697137	14.128409	-6.254511	-4.599490	-1.331696
	MB(ASMA 2003-2017 mean)-MB(PSTW 2017-2019 mean)					MB(ASMA 2017)-MB(PSTW 2017)			
Calibration	13.1409384870107					20.3829202947486			

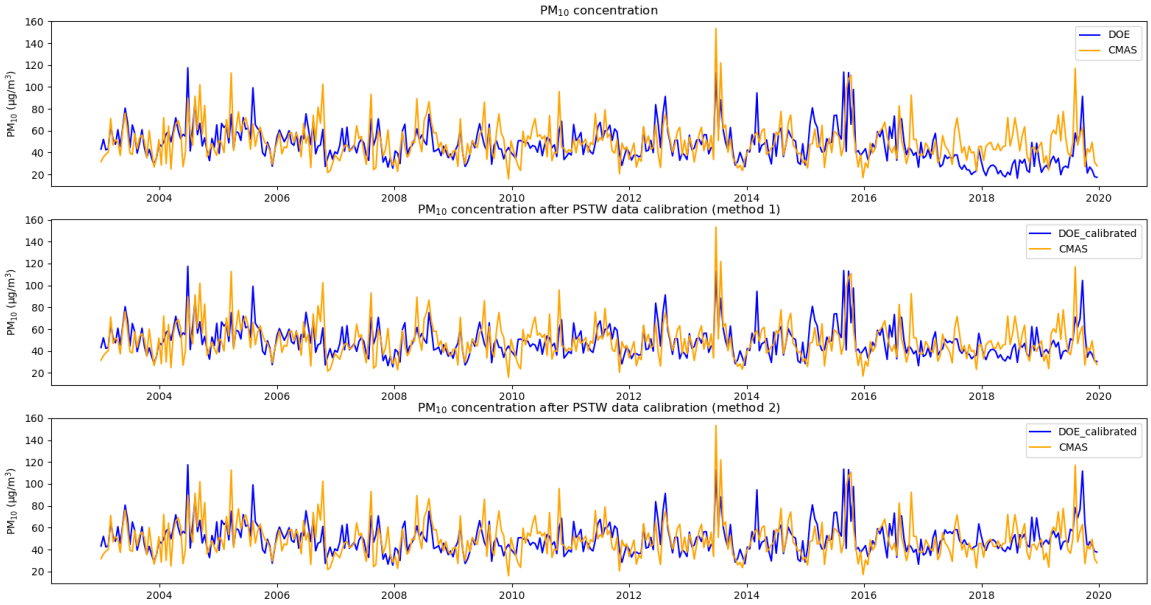


Fig. S3. a) Time series graph of PM_{10} concentration data before calibration in 2003-2019. b) Time series graph of PM_{10} concentration data after calibration in 2017-2019 for the first method. c) Time series graph of PM_{10} concentration data after calibration in 2017-2019 for the second method.

From the figures and tables above, the difference between the two calibration methods can be clearly compared through the presentation of the time series plots. We can see that from the data consistency difference after adjustment, after the second calibration method, the data has less discrepancy with ASMA data, which is more in line with our data needs. Therefore, in this paper, the second calibration method is used to adjust the data, and all subsequent studies are based on the data after calibration using this method.

S2: Names of the measurement station

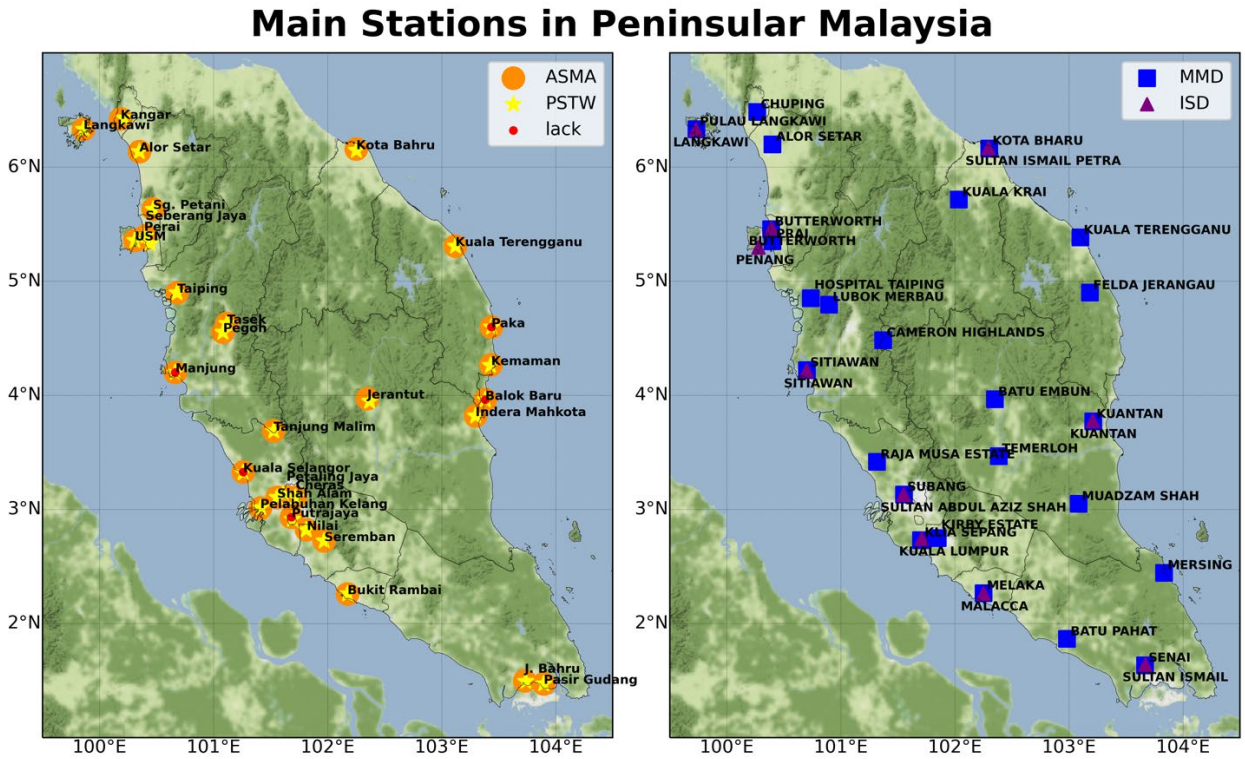


Fig.S4 Map of ASMA/PSTW stations with names (left) and map of MMD/ISD stations with names (right).

Table.S4 Categorization of the regions in PEM for regional CCA analysis

No.	Agglomerations	States
1	Terengganu	Terengganu
2	Johor	Johor
3	Kedah	Kedah, Perlis, Pulau Pinang
4	Kelantan	Kelantan
5	Pahang	Pahang
6	Perak	Perak
7	Selangor (SA)	Selangor, Kuala Lumpur, Putrajaya
8	Negeri Sembilan (NSA)	Negeri Sembilan, Melaka

S3: Shapiro-Wilk test of normality of all variables

In order to use Pearson correlation analysis, first perform a normality test on all the data, the test steps are as follows:

Step 1: Perform Shapiro-Wilk test on the data to check its significance.
Step 2: If it does not show significance (P>0.05), it means that it conforms to the normal distribution, otherwise it means that it does not conform to the normal distribution.

Step 3: Usually, it is difficult to meet the test in real research situations, if the absolute value of the sample kurtosis is less than 10 and the absolute value of the skewness is less than 3, combined with the normal distribution histogram, PP diagram or QQ plot, it can be described as basically conforming to the normal distribution.

Table S5. Shapiro-Wilk test on the yearly data

variable name	sample size	median	average value	standard deviation	Skewness	kurtosis	SW test	KS test
temp mean	20	27.3	27.352	0.269	1.421	2.556	0.847(0.005***)	0.242 (0.164)
pmmean	20	47.928	48.117	1.717	-0.662	1.347	0.915(0.079*)	0.223(0.235)
O3mean	20	0.017	0.017	0.002	1.736	2.825	0.772(0.000***)	0.296(0.047**)
CO mean	20	0.609	0.602	0.039	-2.423	5.602	0.646(0.000***)	0.35(0.011**)
NO2mean	20	0.011	0.01	0	-1.338	0.508	0.768(0.000***)	0.293(0.051*)
SO2mean	20	0.003	0.004	0.001	0.866	-0.311	0.899(0.039**)	0.162 (0.617)
Humidity means	20	77.935	77.766	0.769	-1.639	3.948	0.826(0.002***)	0.26(0.112)
rain mean	20	2464.348	2463.973	216.776	-0.053	-0.991	0.96(0.540)	0.109(0.949)
wind	20	1.791	1.801	0.073	1.305	2.812	0.897(0.036**)	0.111 (0.942)

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively

Table S6. Shapiro-Wilk test on the monthly data

variable name	sample size	median	average value	standard deviation	Skewness	kurtosis	SW test	KS test
Temperature	12	27.559	27.506	0.479	-0.129	-1.359	0.939(0.480)	0.132 (0.967)
PM_{10}	12	48.401	48.351	5.009	-0.309	-0.878	0.942(0.527)	0.157(0.883)
O_3	12	0.018	0.018	0.002	-0.323	0.133	0.972 (0.931)	0.155(0.895)
CO	12	0.604	0.597	0.021	-0.61	-1.177	0.887(0.107)	0.229 (0.486)
NO_2	12	0.011	0.011	0.001	-0.554	-1.071	0.906(0.189)	0.166(0.841)
SO_2	12	0.003	0.003	0	-0.087	-1.618	0.918 (0.268)	0.15 (0.914)
Humidity	12	64.603	63.966	2.191	-0.878	0.452	0.928(0.360)	0.173(0.809)
Precipitation	12	186.879	205.331	66.298	0.833	0.07	0.909(0.204)	0.23(0.480)
Wind speed	12	1.687	1.785	0.22	1.106	-0.141	0.82(0.016**)	0.283(0.242)

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively

Table S7. Shapiro-Wilk test on the daily data

variable name	sample size	median	average value	standard deviation	Skewness	kurtosis	SW test	KS test
Temperature	twenty four	26.645	27.508	2.734	0.417	-1.431	0.88(0.008***)	0.159(0.524)

PM_{10}	twenty four	47.997	48.14	3.503	0.052	-1.363	0.928(0.088*)	0.12 (0.841)
O_3	twenty four	0.012	0.018	0.012	0.622	-1.325	0.829(0.001***)	0.217 (0.181)
CO	twenty four	0.571	0.597	0.17	0.454	-0.983	0.924(0.070*)	0.124(0.814)
NO_2	twenty four	0.01	0.011	0.003	0.551	-1.019	0.898(0.020**)	0.17(0.440)
SO_2	twenty four	0.003	0.003	0	-0.61	-0.954	0.89(0.014**)	0.187(0.328)
Humidity	twenty four	80.983	77.584	10.425	-0.433	-1.471	0.865(0.004***)	0.175(0.405)
Wind speed	twenty four	1.441	1.781	0.807	0.617	-1.284	0.828(0.001***)	0.197(0.271)

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively

Table S5-S7 provide the output results from Shapiro-Wilk test on yearly, monthly and daily data sets respectively. Combined with normality test histograms and the fitting of the cumulative probability (P) of calculation observations and the normal cumulative probability (P) plots, all the normality charts are basically bell-shaped (high in the middle and low at both ends), and the degree of fitting in PP plots are acceptable. Although the data are not absolutely normal, they are basically acceptable as a normal distribution.

S4: canonical correlation analysis with precipitation data

The canonical correlation analysis results with precipitation data are presented in Table S8-S11 as follows.

Table S8. Data sets of climate variables and air quality variables.

Set X	Temperature	Precipitation	Humidity		
Set Y	CO	O_3	SO_2	NO_2	PM_{10}

Table S9. Canonical correlation analysis results.

Canonical variates	Canonical correlation	Eigen values	Wilks	Degree of freedom	F	P
Pair 1	0.995	0.99	0.004	15	13.762	<0.001

Table S10. The Canonical loadings of set Y and set X respectively.

	Y_1
PM_{10}	-0.473
O_3	-0.995
CO	0.796
NO_2	0.020
SO_2	0.763
	X_1
Temperature	-0.998
Humidity	0.908
Precipitation	0.033

It can be seen from the above tables that only the first pairs of canonical variables have passed the significance test, which means the correlation between the first pairs of canonical variables is considered to be significant. The correlation coefficient of the first pair of canonical variables is

0.995. This shows that there is a very close relationship between the set X and the set Y. The subsequent analysis focuses on the first pair of canonical variables and the corresponding canonical loadings. From table S10, the canonical loading is the simple correlation coefficient between the variable and all the variables in the same group. The larger the absolute value of the canonical loadings, the stronger the correlation between the items and the canonical variables.

Table S11. Proportion of variance explained by the first pair of canonical variables.

	Set X	Set Y
X_1	60.743	48.589
Y_1	48.09	60.119

Table S11 shows the proportion of explanations of canonical variables, including the proportion of intra-group explanations and cross explanations, to quantitatively judge the amount of original information contained in canonical variables. The canonical variable X_1 explained 48.589% of the information of the indicators in the set Y, and explained 60.743% of the information of the indicators in the set X. The canonical variable Y_1 explains 60.119% of the information for the indicators in set Y and 48.09% of the information for indicators in set X. The whole picture of this pair of canonical variables is illustrated in figure S4. Canonical variable X_1 and Y_1 can capture the information of 60.743% of meteorological variables and 60.119% of air quality variables respectively. Climate variables can represent 61.224% information of air quality variables, which highly indicate correlations between climate and air quality factors. Interestingly, the canonical loading of precipitation is 0.03, which shows that from the perspective of interannual variation, the impact of precipitation on the air quality factor is not obvious, and the phenomenon of significant and irregular annual variation of precipitation can also confirm this result. The specific impact of precipitation factors on air quality needs to be further analyzed and studied in the future.

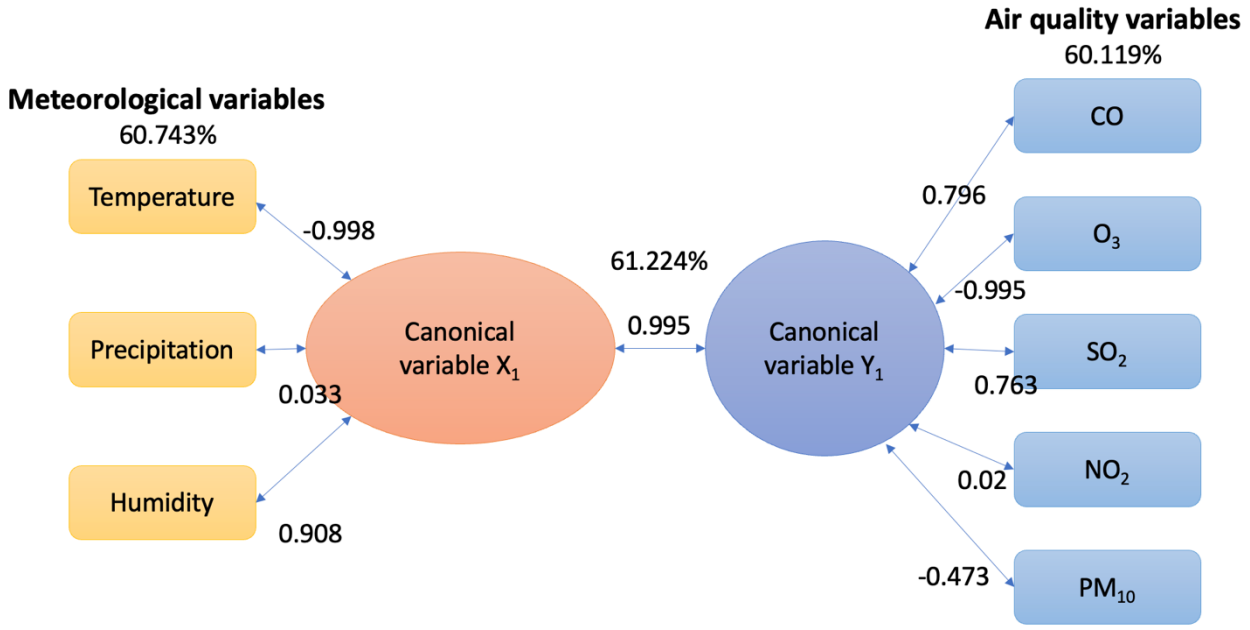


Fig. S5. CCA map of yearly mean temperature, precipitation, relative humidity, and PM_{10}, O_3, CO, NO_2, SO_2 over the observation period.

S5: Regional canonical correlation analysis with precipitation data

Table S12. Regional canonical correlation analysis results including precipitation data.

Regions	Canonical variates	Canonical correlation	Proportion of variance explained	Eigen values	Wilks	Degree of freedom	F	P
Terengganu	Pair 1	0.947	44.90%	0.897	0.015	15	6.779	<0.001
Johor	Pair 1	0.995	50.10%	0.991	0	15	28.281	<0.001
Kedah	Pair 1	0.97	48.83%	0.941	0.009	15	8.35	<0.001
Kelantan	Pair 1	0.97	48.11%	0.941	0.009	15	7.492	<0.001
Pehang	Pair 1	0.936	39.26%	0.876	0.01	15	7.129	<0.001
Perak	Pair 1	0.985	46.43%	0.97	0.003	15	15.204	<0.001
Selangor	Pair 1	0.99	54.90%	0.98	0.005	15	13.032	<0.001
Negeri Sembilan	Pair 1	0.98	58.35%	0.961	0.013	15	8.484	<0.001

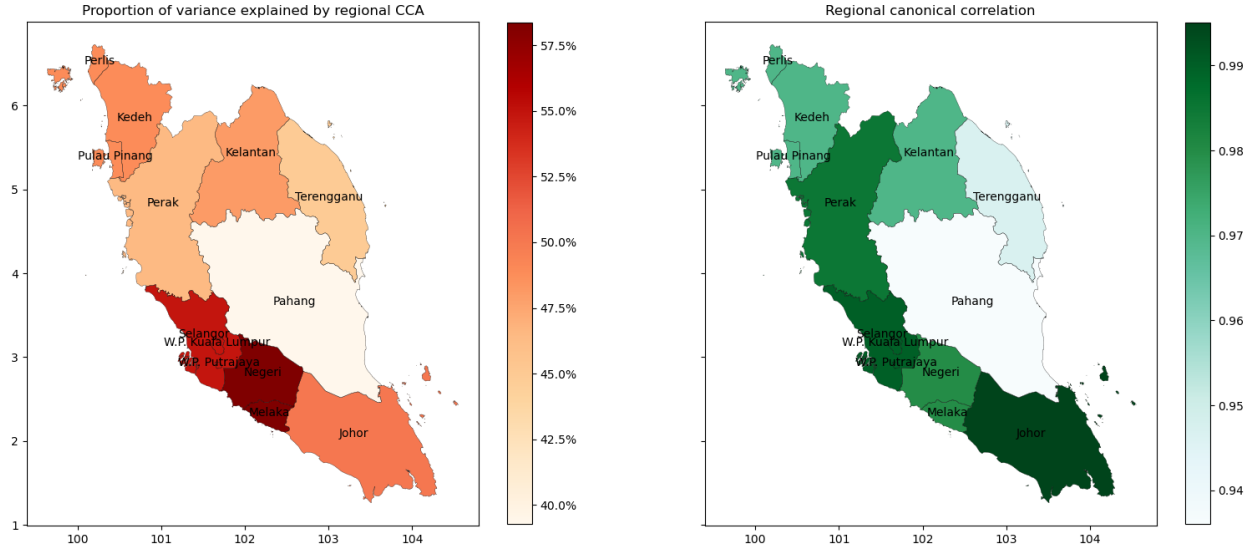
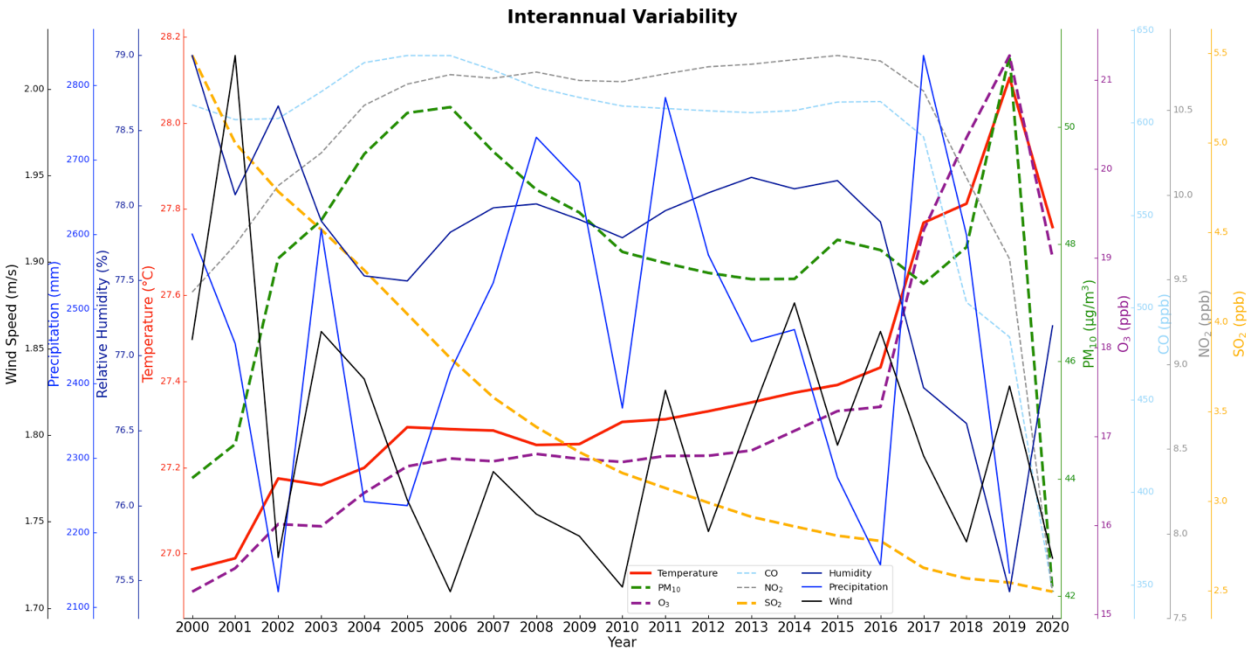


Fig. S6. Proportion of variance explained by regional CCA and regional canonical correlation map of 8 regions in PEM over the observation period.

S6: Reasonings for including year 2020 data

Since 2020 is a very special year, the situation of the Covid-19 runs through the whole year, and both the climate and air quality data have undergone significant turning points as is shown in the figure below. This article wants to focus mainly on the natural variability and the correlation between climate change and air quality rather than the impact of human factors. Therefore, the analysis and discussion of the situation in 2020 is included in the supplementary information rather than in the main article.



From the figure, we can clearly see that due to the impact of the Covid-19, human activities are restricted, and the temperature and pollutants have dropped significantly in 2020, reflecting the huge impact of human activities on climate and air quality. This sudden change affects studies on natural variability, so the analysis of the 2020 data is omitted in the main article.