Archives of Environmental Protection
Vol. 50 no. 4 pp. 116–126

PL ISSN 2083-4772
DOI 10.24425/aep.2024.152901



# Impact of COVID-19 on PM2.5 concentrations in Singapore, Indonesia, and Thailand: Cluster Analysis and Generalized Additive Mixed Models

Wong Ming Wong[1], Shian-Yang Tzeng[2], Hao-Fan Mo[3], Wunhong Su*[4]

[1]Krirk University, Thailand
[2]Quanzhou University of Information Engineering, China
[3]JinWen University of Science and Technology, Taiwan
[4]Hangzhou Dianzi University, China

* Corresponding author's e-mail: whsu@hdu.edu.cn

**Keywords:** COVID-19; PM2.5; Cluster Analysis; Generalized Additive Mixed Model; Thailand, Singapore, Indonesia

**Abstract:** This paper examines the influence of COVID-19-related factors on PM2.5 concentrations (PM2.5) in Singapore, Indonesia, and Thailand from January 2018 to December 2021. Using data from four sources, cluster analysis based on six socioeconomic indices was employed to select these countries for focused analysis. Generalized Additive Mixed Models (GAMM) were applied to assess associations between PM2.5 and COVID-19 factors, including new cases, deaths, vaccinations, stringency index, time series (STOL), and COVID-19 status (dummy variable). Results show that PM2.5 levels in Singapore and Indonesia were significantly impacted by COVID-19 measures, with F-statistics for new cases (22.875, $p < 0.001$), deaths (12.563, $p = 0.012$), as well as significant associations for vaccinations ($t = 5.976$, $p < 0.001$), stringency index ($t = 5.124$, $p < 0.001$), and the dummy variable ($t = 6.624$, $p < 0.001$). In contrast, PM2.5 levels in Thailand were unaffected by these factors, likely due to seasonal pollution sources. The model explains 90.3% of the variation in PM2.5 (adjusted $R^2 = 0.872$).

This paper offers important insights for policymakers on incorporating air quality into health policies and highlights how pandemic responses varied across countries. By examining the impact of COVID-19 factors on PM2.5 in different nations, the study enhances understanding through detailed data and averaging periods. It reveals differences in how countries' air quality responded to the pandemic, contributing to discussions on environmental management and public health. These findings inform policy decisions and facilitate discussions on better managing environmental and health challenges during global crises.

## Introduction

The COVID-19 pandemic (COVID-19) has impacts that extend beyond public health, influencing environmental parameters such as air quality, while also altering consumer behaviors and business models globally. This paper focuses on PM2.5 concentrations (PM2.5) and examines their fluctuations in relation to COVID-19-related factors in Southeast Asia. Previous studies have highlighted shifts in consumer behaviors, such as increased reliance on online purchasing and delivery services driven by health concerns and lockdown measures (Li et al. 2020; Lim et al. 2021; Wang et al. 2020; Webster, 2020). These studies can be categorized into three themes: public health, consumer behavior, and environmental outcomes. By investigating how these factors and regional policies influence PM2.5, this paper contributes a detailed analysis that enhances understanding of the pandemic's multifaceted effects within the Southeast Asian context.

The first theme explores how COVID-19 has impacted consumer behavior, such as purchasing decisions (Li et al. 2020; Wang et al. 2020). The second theme addresses public health concerns, such as social distancing and diseases related to COVID-19 (Barouki et al. 2021; Liao et al. 2020). The third theme focuses on environmental and health-related issues, such as the relationship between air pollution, COVID-19 vaccination, and daily confirmed COVID-19 cases (Lim et al. 2021; Meo et al. 2021; Wetchayont, 2021). This theme also considers the impact of pollution, particularly air quality, on public health outcomes (Frączek et al. 2023; Godłowska et al. 2022).

The third theme specifically focuses on the association between COVID-19 and air pollution. Numerous studies are currently being conducted to examine the association between COVID-19 and meteorological parameters such as climate, air pollution, and the environment (Kotsiou et al. 2021; Liu et al. 2021; Lorenzo et al. 2021; Meo et al. 2021; Valdés Salgado et al. 2021; Wetchayont, 2021). For instance, Liu et al. (2021) examine the effect of air pollution on COVID-19 in China, Japan, Korea, Canada, the United States, Russia, England, Germany, and France, finding that high levels of air pollution combined with a reduction in air quality contribute to an increase in new confirmed cases of COVID-19. Kotsiou et al. (2021) reported that PM2.5 negatively influences new COVID-19 cases in Italy. However, Valdés Salgado et al. (2021) found no statistically significant link between COVID-19 death rates and PM2.5 or PM10 in Chile.

On the other hand, numerous studies have compared air pollution levels before and after the COVID-19 lockdown. These studies generally conclude that air pollution, particularly PM2.5, decreased during the COVID-19 lockdown due to reduced transportation emissions and decreased business and industrial activities (Gkatzelis et al. 2021; Kaewrat & Janta, 2021; Lee & Finerman, 2021; Wetchayont, 2021). For instance, Lee and Finerman (2021) found a 10% to 20% reduction in air pollution emissions in South Korea during the COVID-19 pandemic. They noted that a 1% reduction in commuting flows led to a corresponding reduction in air pollutants – approximately 0.08% to 0.17 % for air pollutants such as PM2.5, PM10, $NO_2$, CO, and $SO_2$. Similarly, Wetchayont (2021) examined the impact of the COVID-19 lockdown on air pollution in Greater Bangkok, Thailand, across three time periods: (1) before, (2) during, and (3) after the lockdown. The COVID-19 lockdown restricted certain types of traffic as well as commercial and industrial activities, resulting in significant changes to the levels and characteristics of air pollution (Wetchayont, 2021).

The theme of air quality is significantly influenced by external environmental factors and localized pollution sources, as discussed in the study by Frączek et al. (2023) and Godłowska et al. (2023). Frączek et al. identified shortcomings in natural ventilation systems within schools, linking them to inconsistent fungal aerosol concentrations influenced by factors such as relative humidity and inadequate air exchange. Conversely, Godłowska et al. (2022) demonstrated how urban air quality forecasts can be improved using precise meteorological data and detailed emission inventories, emphasizing the positive impact of Krakow's policy to phase out coal furnaces. Furthermore, Wong et al. (2024) employed cluster analysis and Generalized Additive Mixed Models (GAMM) to study the effects of COVID-19 and population dynamics on the Air Quality Index (AQI) in Malaysian cities. Their findings indicate that while population density significantly impacts AQI, the COVID-19 pandemic did not alter AQI variations, suggesting an anticipated overall improvement in air quality.

### *Research Objective and Questions*

Previous studies on the association between COVID-19 and air pollution have demonstrated that pollutants such as PM2.5, $NO_2$, $SO_2$, alongside overall air quality, can affect COVID-19 infection rates (Lee & Finerman, 2021; Lim et al. 2021; Liu et al. 2021; Meo et al. 2021; Frączek et al. 2023; Godłowska et al. 2022; Wong et al. 2024). However, these studies are often limited by their narrow focus on a single country and the use of data collection periods shorter than 36 months. Moreover, while many of these studies overlook issues related to time-series data, Wong et al. (2024) address these concerns, though their study is also limited to a single country context.

Consequently, this paper aims to address the research gap by concentrating on the relationship between PM2.5 with COVID-19-related factors in Southeast Asia. However, a key limitation of this research lies in the use of secondary data, which often presents challenges such as non-parametric, and nonlinear characteristics, as well as missing values. Thus, the study employs Cluster Analysis and GAMM to meet the research gap (Augustin et al. 2009; Chen, 2000; Wong et al. 2024; Wood, 2006, 2011). The central research question is: Does an association exist between COVID-19-related factors, including COVID-19 new cases, deaths, vaccinations, stringency index, time series data, and COVID-19 status (presence or absence of cases), and PM2.5 in various countries?

### *Generalized Additive Mixed Models*

Generalized Additive Mixed Models (GAMM) are advanced statistical methods that enhance Generalized Additive Models (GAM) by including random effects, allowing for greater flexibility in handling hierarchical or correlated data structures. Built on the framework of Generalized Linear Models (GLM), GAMM uses smooth functions to model nonlinear association between variables. These models are particularly effective for analyzing time-series data and other datasets with complex dependencies (Wood, 2006, 2011, 2017).

The GAM is a smooth fitting curve consisting of polynomials like a spline, designed to approximate the data as closely as possible without causing excessive overfitting (Brömssen, 2016; Constantinescu, 2019; Hastie & Tibshirani, 2017; Wood, 2006, 2011, 2017). One key advantage of GAM is its ability to explore nonlinear and nonmonotonic correlations between variables. The general equation for GAM is provided below,

$$y = \alpha + f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) + \varepsilon \qquad (1)$$

In Equation (1), α represents the intercept, f denotes the smooth function, $y$ is the response variable, $\chi_i$ is the explanatory variable $\forall i=1,2,\cdots,n$, and ε is the error term (Brömssen, 2016; Fahrmeir & Lang, 2001; Hastie & Tibshirani, 2017; Wood, 2017).

GAM employs a smoothing function technique, developed by Trevor Hastie and Robert Tibshirani, to model the relationship between a dependent variable, Y, and a set of predictor variables, X (Hastie & Tibshirani, 2017). The model uses a contiguous function to link these variables, addressing linear models that struggle with dependent variables. Numerous regression issues require flexible, semi-parametric predictors to capture covariate associations, often incorporating unobserved random effects (Chen, 2000; Fahrmeir & Lang, 2001; Wood, 2006, 2011; Wood, 2017).

GAM assumes data independence, however, time series data are inherently dependent. Despite this, the GAMM is widely used for time series, spatial, hierarchical, clustering, or nested design models (Brömssen, 2016; Constantinescu, 2019; Fahrmeir & Lang, 2001; Groll & Tutz, 2012; Hastie

& Tibshirani, 2017; Wood, 2017). The generalized additive mixed model (GAMM) can be expressed as (Wood, 2017):

$$\mathcal{Y}_{\mathrm{i}} = \mathrm{X}_i\beta + f_1(\mathcal{X}_{1\mathrm{i}}) + f_2(\mathcal{X}_{2\mathrm{i}}, \mathcal{X}_{3\mathrm{i}}) + \cdots + \mathrm{Z}_i\mathrm{b} + \epsilon_i \qquad (2)$$

In equation (2), where $\mathcal{Y}_{\mathrm{i}}$ is a univariate response, $X_i$is the fixed effects design row, and $f_k(\cdot)$ are smooth functions of the covariates $X_{ki}$. The term $Z_i$is the design row for random effects, and b is a vector of random effect coefficients assumed to follow a multivariate normal distribution:

$$\mathrm{b} \sim \mathrm{N}(0, \psi),$$

where ψ is an unknown positive definite covariance matrix. The residual errors, $\varepsilon_i$, are also assumed to be normally distributed:

$$\epsilon \sim \mathrm{N}(0, \Lambda),$$

with Λ being the residual covariance matrix, often assumed to have a simple structure.

A smooth function f(x) can be represented as a combination of basis functions, $b_j$ (X) with coefficients $\beta_j$.

$$f(\mathrm{X}) = \sum_{j=1}^{J} b_j\,(\mathrm{X})\beta_j \qquad (3)$$

A smoothness penalty controls the “wiggliness” or complexity of the smooth function:

$$J(f)= \beta^T\,\mathrm{S}\beta$$

S is a positive semi-definite matrix. The penalty helps maintain a balance between fitting the data and ensuring the function remains smooth.

In addition, GAMM utilizes a penalized regression method in machine learning, which creates smoothness for multiple variables. This approach offers several benefits: (1) Rescaling covariates linearly is invariant; (2) The range of smoothness is adjustable; (3) Reduced smoothness improves computing efficiency; (4) Smoothing can be automatically generated from any edge smoothing base, allowing for versatile modeling; (5) Smoothing can be easily incorporated into standard linear or generalized linear mixed models (Constantinescu, 2019; Fahrmeir & Lang, 2001; Groll & Tutz, 2012; Tuerlinckx et al. 2006; Wood, 2006; Wood, 2017). Furthermore, GAMM processes nonlinear relationships between explanatory and dependent variables by adapting machine learning techniques to modeling. It combines data mining characteristics with enhanced descriptive power (Constantinescu, 2019; Fahrmeir & Lang, 2001; Groll & Tutz, 2012; Wood, 2017).

GAMM has numerous applications in the fields of medicine, finance, public health, business, and management. For instance, Augustin et al. (2009) found that the defoliation time of tree canopy in Baden-Württemberg, Germany, depends on its location and the level of air pollution. The data on air pollution includes various crossover and time-dependent features, including repeated measurements.

Due to GAMM’s ability to combine the benefits of both the GAM and the generalized linear mixed model (Constantinescu, 2019; Fahrmeir & Lang, 2001; Groll & Tutz, 2012; Hastie & Tibshirani, 2017; Tuerlinckx et al. 2006; Wood, 2006, 2011; Wood, 2017), GAMM can conduct data mining to uncover complex relationships among variables. Therefore, this paper examines how GAMM can analyze air pollution datasets through data mining, detect variables (such as COVID-19-related factors), and examine changes in time series data across two categories: countries and PM2.5.

## Materials and methods

### Research Design

The research methodology of this paper is divided into two steps. First, a cluster analysis was performed on the six-core socioeconomic indexes of ten Southeast Asian countries using SPSS to identify homogenous groups based on shared characteristics. Second, the paper unitized GAMM via R Studio to analyze the association between PM2.5 concentrations and COVID-19-related factors. This two-step approach enables a focused examination of the variables significantly influencing air quality within the selected cluster of Southeast Asian countries.

### Data Extraction

This paper accessed four websites to download the relevant data. First, the “Basic Statistics 2020” dataset was downloaded from the Asian Development Bank’s website (Asian Development Bank, 2020). This dataset provides the core socioeconomic indexes for ten Southeast Asian countries. The Asian Development Bank defines these six core indexes as follows: (1) total population (in millions), (2) average annual population growth rate, (3) nominal gross domestic product (GDP) (in billions of USD), (4) annual GDP growth rate, (5) inflation rate, and (6) current account balance (as a percentage of GDP). These six variables were utilized in the first step of the cluster analysis.

This paper obtained data on COVID-19-related factors from the “Our World in Statistics” website (Mathieu et al. 2021). In the second step, GAMM analysis was used to analyze six variables: (1) COVID-19 new cases, (2) COVID-19 deaths, (3) COVID-19 vaccinations, (4) stringency index, (5) time series STOL, and (6) COVID-19 status. Specifically, the time series variable (STOL) represents monthly data from January 2018 to December 2021, The COVID-19 status variable is a binary indicator (‘Dummy’), where 0 represents no COVID-19 cases, and 1 represents the presence of COVID-19 cases, based on the same time series. However, the dataset is limited by country availability. For instance, Vietnam, Laos, and Myanmar lack complete datasets on COVID-19-related factors.

The unit for COVID-19-related factors is a numerical count, while PM2.5 is measured in micrograms per cubic meter (μg/m³). This paper uses data from two sources that provide country-level PM2.5 data. Specifically, PM2.5 data for Thailand and Indonesia were obtained from the ‘Berkeley Earth’ website (Berkeley Earth, 2020), and data for Singapore were sourced from the ‘World Air Quality Project’ website (The World Air Quality Index Project, 2022). The stringency index is a composite measure scaled from 0 to 100, with 100 representing the strictest response. It is calculated based on nine indicators, including school closures, workplace closures, and travel prohibitions, as reported by the ‘Our World in Statistics’ website (Mathieu et al. 2021).

### Process

This paper gathered three datasets from four sources for a two-step analysis. In the first step, six variables representing the core socioeconomic indexes of ten ASEAN countries were analyzed using cluster analysis. This analysis identified a cluster of countries with similar socioeconomic backgrounds from ten ASEAN countries, which were selected for further examination.

**Table 1.** *The Descriptive Statistics of Ten Southeast Asian Countries' Core Indexes*

| Index | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|
| Total Population (Million) | 0.452 | 266.912 | 65.562 | 80.461 |
| Ave. Annual Population Growth Rate | 0.308 | 2.069 | 1.114 | 0.484 |
| Nominal GDP ($ Billion) | 13.469 | 1119.152 | 314.898 | 338.577 |
| Annual Growth Rate of GDP | 0.733 | 7.054 | 4.802 | 2.072 |
| Inflation Rate | -0.411 | 8.626 | 2.351 | 2.523 |
| Current Account Balance (% of GDP) | -17.595 | 16.970 | 0.754 | 9.790 |

Note: N = 10 (Brunei, Cambodia, Indonesia, Lao. Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam)

Second, the COVID-19-related factors collected in this paper include (1) new COVID-19 cases, (2) COVID-19 deaths, (3) COVID-19 vaccinations, and (4) the stringency index for the period from 2018 to 2021. These data are reported daily; therefore, the paper used measures of central tendency to calculate monthly averaged values of all variables. Additionally, the paper designed two derived variables: (1) 'STOL,' which represents an ordered monthly time series from January 2018 to December 2021), and (2) 'Dummy,' a binary variable indicating whether a COVID-19 case occured in a given month (0 = no case, 1 = case). As a result, six variables represent parameters associated with COVID-19-related factors.

Third, PM2.5 values, measured in micrograms per cubic meter (μg/m³), and recorded hourly. Similar to the COVID-19-related factors, measures of central tendency were applied to calculate monthly averages for PM2.5 values from January 2018 to December 2021. Finally, the sample size for both COVID-19-related factors and PM 2.5 is 144, representing a cluster country over 48 months.

All original data in this paper were averaged monthly using a central tendency approach, specifically the arithmetic mean, to calculate yearly values. This method was chosen because GAMM operates most effectively with central tendency measures, which represent a significant or typical value for a dataset, typically the mean (Hastie & Tibshirani, 2017; Tuerlinckx et al. 2006; Weisberg, 1992; Wood, 2017). The arithmetic mean was favored because it effectively aggregates all data points into a single representative value.

Considering the objectives and strengths of GAMM, this paper adopts central tendency measures for analyzing COVID-19-related factors and PM2.5 data. This approach was selected to address two primary challenges presented by the dataset. First, the secondary data exhibit characteristics such as non-parametric, nonlinear distributions, as well as missing values. Second, there is uneven data availability, particularly

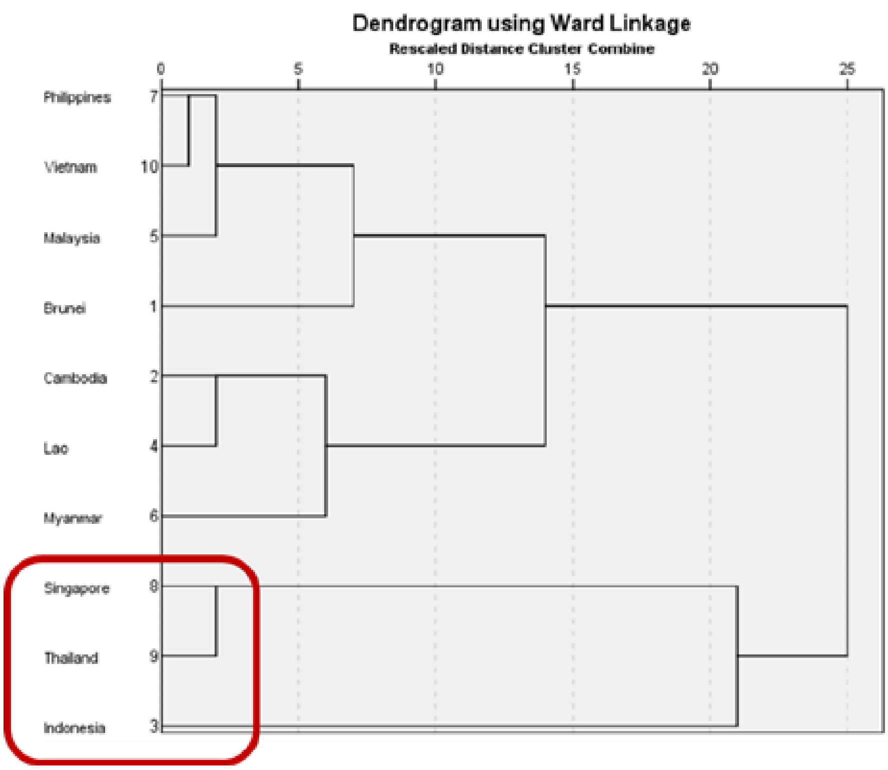


**Fig 1.** The Cluster Analysis of Ten ASEAN Countries

from smaller regions, such as certain cities in Indonesia, where complete datasets may not be accessible. These issues necessitate a statistical method capable of effectively handling such complexities.

Furthermore, past studies have adopted different analysis concepts, ranging from techniques to data. For example, Mathieu et al. (2021) showed that, although cities have varying numbers of new COVID-19 cases, they applied central tendency measures to present their analysis results. Lee and Finerman (2021) also used central tendency measures of PM2.5, PM10, $NO_2$, CO, and $SO_2$ to represent South Korea in their analysis of the association between air pollution emissions and the country. Wong et al. (2024) similarly employed central tendency measures to analyze COVID-19 and AQI data in Malaysian cities using GAMM. Therefore, each country's COVID-19-related factors and PM2.5 data can be combined into a single index based on central tendency.

## Data analysis

### *Descriptive Statistics of Southeast Asian Countries' Core Six Indexed*

The Asian Development Bank provides six core indexes to describe the socioeconomic characteristics of countries. Table 1 presents the descriptive statistics for these core indexes across ten Southeast Asian countries.

### *Cluster Analysis*

The first step involved conducting a cluster analysis using SPSS, applying six core indexes to group ten Southeast Asian countries, as shown in Figure 1. This analysis categorized the countries into three distinct clusters: Cluster One includes the Philippines, Vietnam, Malaysia, and Brunei; Cluster Two comprises Cambodia, Laos, and Myanmar; and Cluster Three consists of Singapore, Thailand, and Indonesia. Each cluster represents countries with shared demographic and economic characteristics. This structured approach facilitates analysis of patterns among Southeast Asian countries, enabling an organized exploration of similarities and differences within the region.

Cluster Three was selected for the next step of analysis because of its comprehensive COVID-19-related data and PM2.5 measurements. In contrast, several countries in Clusters One and Two lack complete datasets for these variables. Singapore, Thailand, and Indonesia, however, provide full datasets, enabling a thorough examination of the relationship between COVID-19-related factors and PM2.5. This selection supports the paper's bojective of exploring the association between the pandemic and air quality issues. Consequently, Cluster Three is well-suited for analyzing the pandemic and pollution in Southeast Asia.

### *Descriptive Statistics of COVID-19 Related Factors and PM2.5*

The data extraction section details the numerical counts of COVID-19-related factors, averaged monthly, which include new cases, deaths, vaccinations, and the stringency index. The stringency index, a composite measure ranging from 0 to 100, is derived from nine indicators such as school closures, workplace closures, and travel restrictions, with 100 indicating the strictest response (Mathieu et al. 2021). PM2.5levels, expressed in micrograms per cubic meter (μg/m³), were also analyzed. Table 2 summarizes the descriptive statistics for

**Table 2.** The Descriptive Statistics of COVID-19 & PM2.5 in Thailand, Singapore, and Indonesia

| Country | Factor | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Thailand N = 48 | New Cases | 0.000 | 607442.000 | 42758.288 | 118475.997 |
| | Deaths | 0.000 | 6732.000 | 417.269 | 1266.713 |
| | Vaccinations | 0.000 | 19468115.000 | 1533918.577 | 4357256.325 |
| | Stringency Index | 0.000 | 76.264 | 23.240 | 28.208 |
| | PM 2.5 | 8.595 | 46.945 | 24.532 | 11.972 |
| Singapore N = 48 | New Cases | 0.000 | 101853.000 | 5373.173 | 17205.540 |
| | Deaths | 0.000 | 312.000 | 15.923 | 61.859 |
| | Vaccinations | 0.000 | 2143437.000 | 224605.462 | 486526.535 |
| | Stringency Index | 0.000 | 74.642 | 22.819 | 25.899 |
| | PM 2.5 | 34.806 | 98.400 | 50.769 | 9.954 |
| Indonesia N =48 | New Cases | 0.000 | 1231386.000 | 81975.385 | 204133.225 |
| | Deaths | 0.000 | 38904.000 | 2771.038 | 7366.514 |
| | Vaccinations | 0.000 | 51168870.000 | 5239050.308 | 12828159.213 |
| | Stringency Index | 0.000 | 73.822 | 27.709 | 33.011 |
| | PM 2.5 | 0.000 | 20.505 | 9.043 | 5.269 |

these factors across Thailand, Singapore, and Indonesia, including the minimum, maximum, mean, and standard values. This provides a quantitative overview of COVID-19-related factors and PM2.5 conditions in these countries.

Figures 2 to 6 present the variables analyzed for Thailand, Singapore, and Indonesia. Each figure focuses on one variable: COVID-19 new cases, COVID-19 deaths, COVID-19 vaccinations, the stringency index, or PM2.5 levels in each country. These variables are tracked over the period from January 2018 to December 2021. The figures facilitate a comparison of temporal changes in each variable within each country. By examining these trends, the paper underscores the impact of COVID-19 and environmental factors on the studied regions.

The figures illustrate the trends and fluctuations of each variable over the study period. For instance, Figure 2 shows that the number of new COVID-19 cases in Thailand rose steadily until August 2021, peaking before declining. Figure 6 indicates missing data for Indonesia between March and August 2018, which may affect the analysis for that period. Additionally, the stringency index varies among the countries, reflecting differences in government responses to COVID-19. PM2.5 levels exhibit fluctuations in air quality over time, offering insights into environmental impacts. These detailed visual representations provide a comprehensive understanding of the temporal dynamics affecting Thailand, Singapore, and Indonesia.

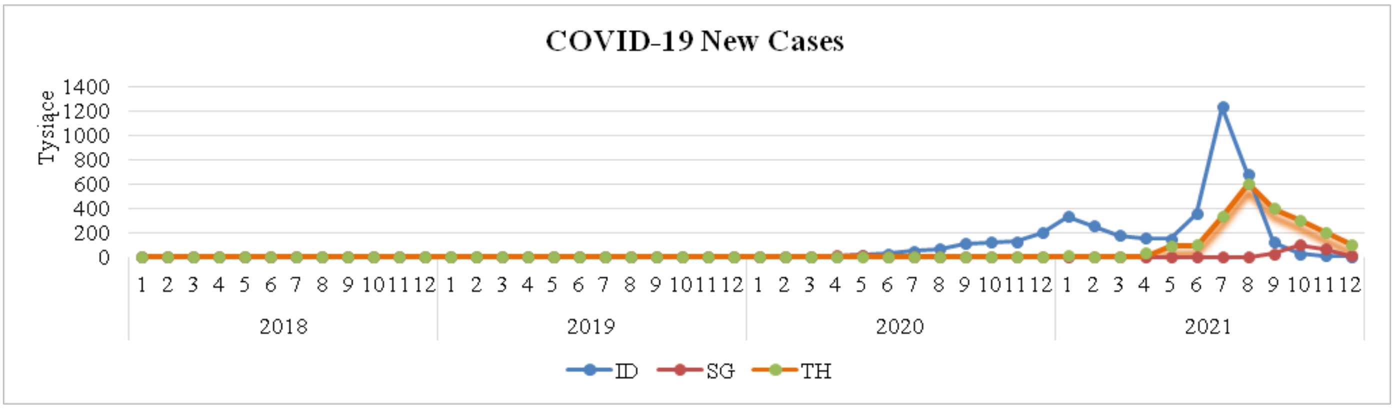


**Fig 2.** Monthly COVID-19 New Cases in Indonesia, Singapore, and Thailand (January 2020 to December 2021)

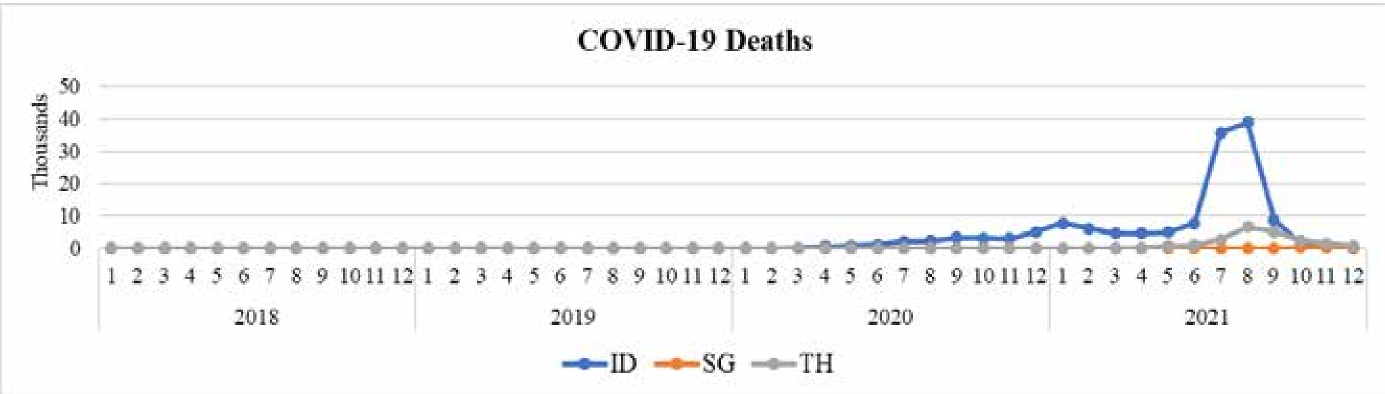


**Fig 3.** Monthly COVID-19 Deaths in Indonesia, Singapore, and Thailand (January 2020 to December 2021)

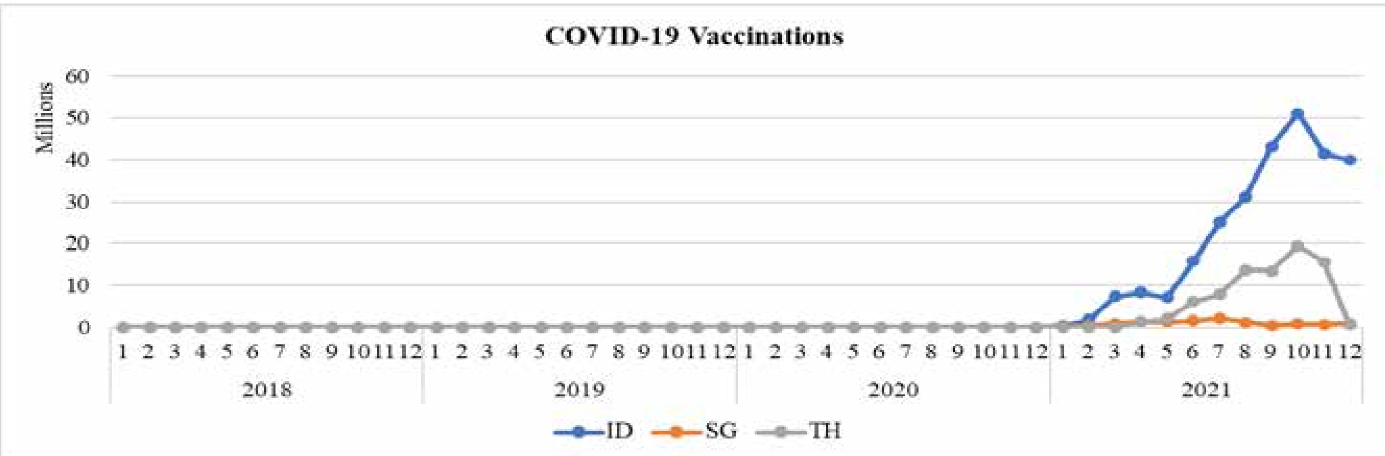


**Fig 4.** Monthly COVID-19 Vaccinations in Indonesia, Singapore, and Thailand (January 2020 to December 2021)

Figures 2 to 6 depict trends in COVID-19 cases, deaths, vaccinations, the stringency index, and PM2.5 levels for Indonesia, Singapore, and Thailand. Figure 2 shows a sharp rise in monthly COVID-19 new cases starting in early 2020, peaking in mid-2021. Indonesia experienced the highest peak, followed by Thailand, while Singapore reported significantly lower case numbers. Figure 3 shows a similar trend in deaths, with Indonesia recording the highest mortality, and Singapore and Thailand reporting comparatively fewer deaths. Figure 4 highlights the rise in vaccinations, with Indonesia administering the most, followed by Thailand, while vaccination levels stabilized across all three countries by late 2021. Figure 5 describes the stringency index, which remained elevated throughout 2020 and 2021. Singapore and Thailand exhibited similar levels initially, while Indonesia maintained moderately stable restrictions. Finally, Figure 6 presents PM2.5, trends, with Singapore reporting the highest concentrations and Indonesia the lowest. PM2.5 levels fluctuated over time but showed no consistent upward or downward trend.

### *Generalized Additive Mixed Model*

The Generalized Additive Mixed Model (GAMM) was established using the R Studio's GAMM package (gamm4 version 0.2-6) to examine the impact of selected variables on PM2.5. The model equation is structured as follows:

$$\text{Equation: PM2.5} \sim \text{Nation} + s(\text{STOL}) + s(\text{New Cases}) + s(\text{Deaths}) + \text{Vaccinations} + \text{Stringency Index} + \text{Dummy} \quad (4)$$

In Equation (4), 'Nation' represents country indicators for Thailand, Singapore, and Indonesia. Within the GAMM framework, 'Nation' is treated as a categorical variable with three levels corresponding to the three countries. Typically, one country serves as the reference category, in this case, Thailand, against which the other countries are compared. The coefficients for 'Nation' in the model output represent the differences in PM2.5 levels between the reference country and the other countries.

The smooth terms s(STOL), s(New Cases), and s(Deaths) capture the nonlinear associations between these variables and PM2.5 levels. STOL represents the monthly time series, while New Cases and Deaths are COVID-19-related factors. Vaccinations and the Stringency Index are included as linear terms, and Dummy is a binary variable indicating the presence or absence of COVID-19 during the study period.

The GAMM analysis generated two outputs, summarized in Tables 3 and 4, detailing the effects of six variables based on

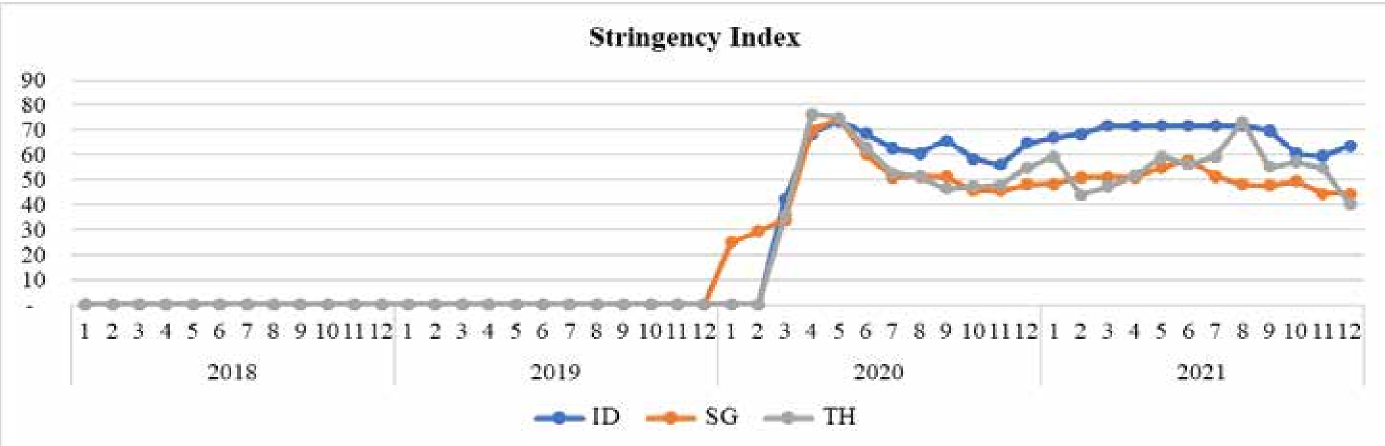


**Fig 5.** Monthly Stringency Index in Indonesia, Singapore, and Thailand (January 2020 to December 2021)

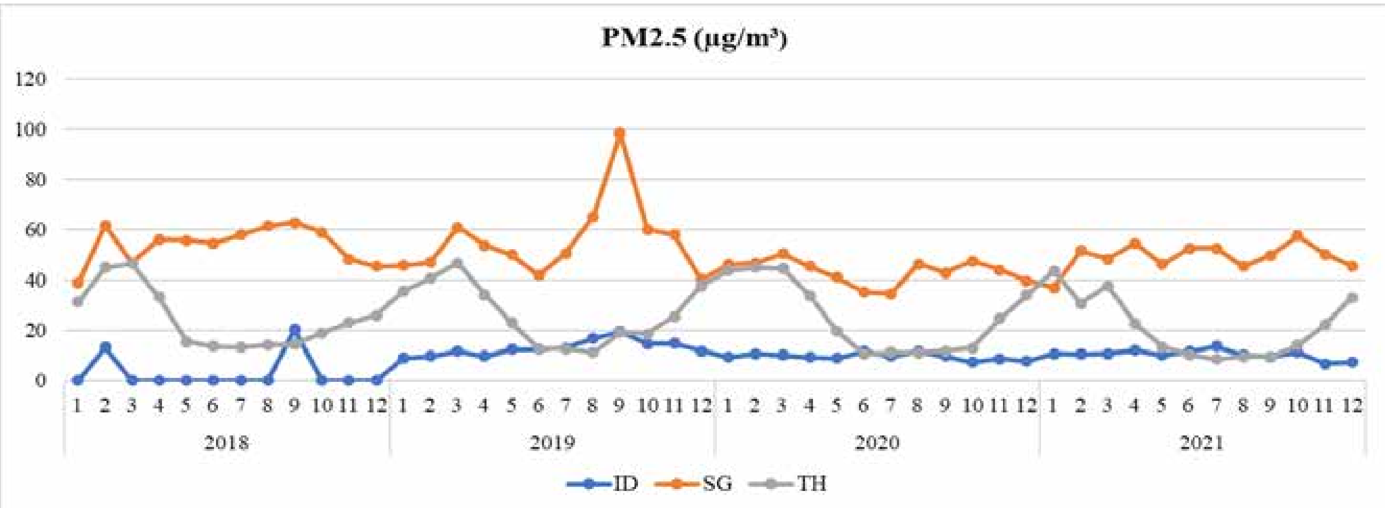


**Fig 6.** PM2.5 in Indonesia, Singapore, and Thailand (January 2020 to December 2021)

**Table 3.** GAMM's Fixed-Effect Model on PM2.5

| Variable | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 5.317 | 0.803 | 6.624 | 0.000*** |
| Singapore | 17.150 | 1.814 | 9.455 | 0.000*** |
| Thailand | 0.131 | 2.516 | 0.052 | 0.959 |
| Vaccinations | 0.000 | 0.000 | 5.976 | 0.000*** |
| Stringency Index | 0.290 | 0.057 | 5.124 | 0.000*** |
| Dummy | 5.317 | 0.803 | 6.624 | 0.000*** |

Note: *** p< 0.001; ** p<0.01; * p< 0.05

**Table 4.** GAMM's Smoothing Fixed-Effect Model on PM 2.5

| Variable | edf | Ref. df | F | p-value |
|---|---|---|---|---|
| s(STOL) | 3.008 | 3.490 | 9.672 | 0.002** |
| s(New Cases) | 1.755 | 2.085 | 22.875 | 0.000*** |
| s(Deaths) | 1.165 | 1.184 | 12.563 | 0.012* |

Note: *** p< 0.001; ** p<0.01; * p< 0.05;
R-sq.(adj) = 0.872 Deviance explained = 90.3%

the model equation. Table 3 presents the fixed-effect estimates, capturing the non-smoothing impacts within the GAMM framework. Table 4 highlights the smoothing effects of STOL, New Cases, and Deaths on PM2.5 levels.

Table 3 presents the fixed-effect estimates from the GAMM model. Each column in Table 3 conveys specific information: the 'Estimate' column shows the predicted effect of each variable on PM2.5 levels, 'Std. Error' indicates the standard error associated with these estimates, 't value' provides the t-statistic for each effect, and 'Pr(>|t|)' represents the p-value, reflecting statistical significance.

Table 3 shows the factors influencing PM2.5 as estimated by the GAMM fixed-effect model. The estimate for Singapore (17.150) indicates that PM2.5 levels are higher in Singapore than in Indonesia when other factors are held constant. Both the vaccination rate and the stringency index have significant positive associations with PM2.5, suggesting that increases in these variables correspond to higher PM2.5 levels. In contrast, the variable for Thailand does not significantly affect PM2.5, suggesting that PM2.5 levels in Thailand are not impacted by the factors included in this model. The Dummy variable is also significant, reflecting baseline differences in PM2.5 across the dataset.

Table 4 presents the smoothing effects of STOL, New Cases, and Death on PM2.5 levels. Each column conveys specific information: 'edf' (effective degrees of freedom) measures the complexity of each smooth term, 'Ref. df' indicates the reference degrees of freedom, 'F' shows the F-statistic for each smoothing effect, and 'p-value' reflects the significance level.

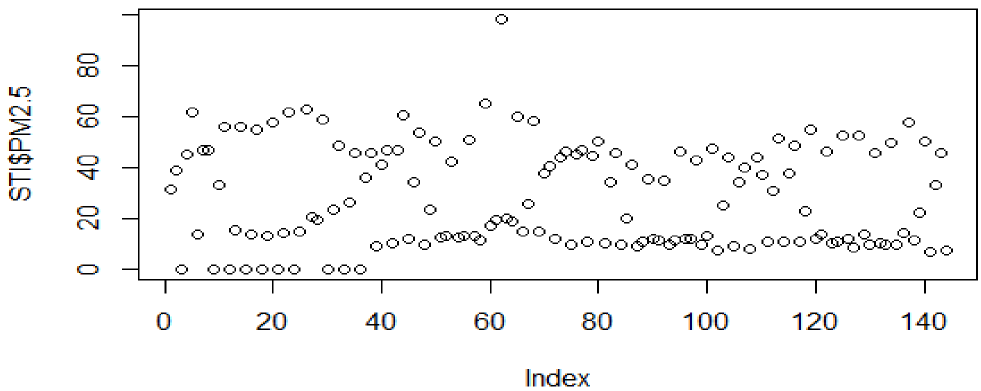


**Fig 7.** Smoothing Fixed-Effect Model on PM2.5

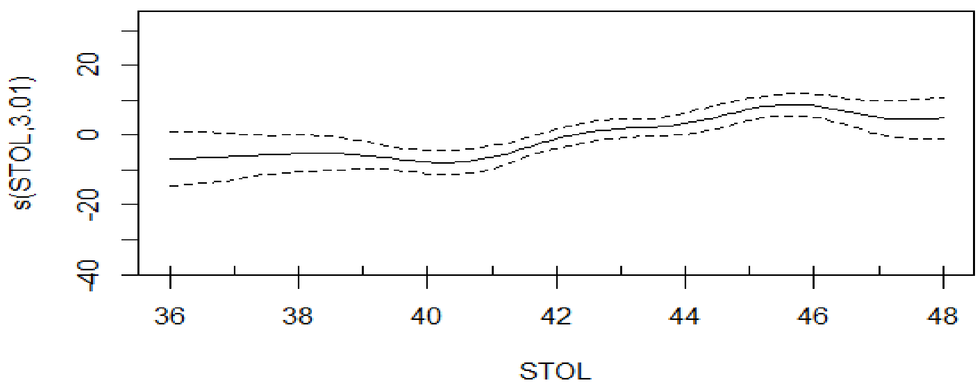


**Fig 8.** Smoothing Fixed-Effect Model on Time Series (January 2018 to December 2021)

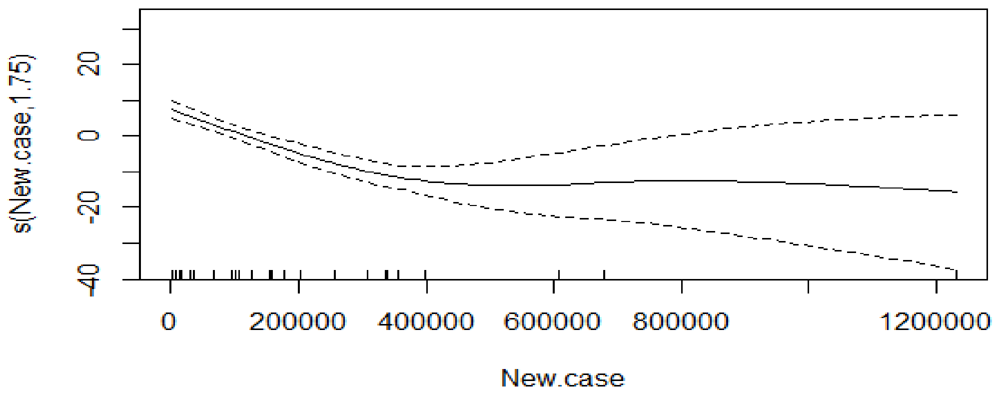


**Fig 9.** Smoothing Fixed-Effect Model on COVID-19 New Cases

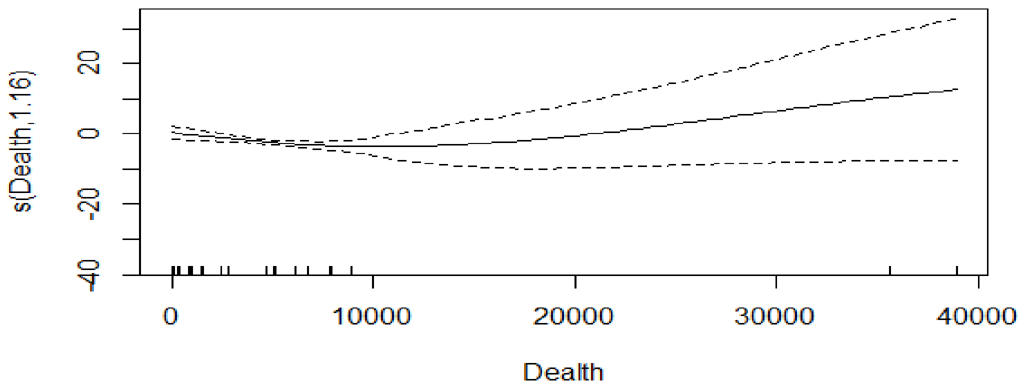


**Fig 10.** Smoothing Fixed-Effect Model on COVID-19 Deaths

The results indicate that the time series (STOL), New Cases, Deaths, and COVID-19 status (Dummy) are significantly associated with PM2.5, with varying degrees of impact as indicated by their respective F-statistics. The model's performance is assessed using an adjusted R-squared value of 0.827, demonstrating that the model accounts for 82.7% of the variation in PM2.5 levels. Additionally, the model explains 90.3% of the deviance in PM2.5, suggesting a high level of fit for the included variables.

Based on the GAMM results shown in Figures 7 to 10, the smoothing effects of various factors on PM2.5 are presented. Figure 7 depicts the GAMM smoothing effect on PM2.5 over time, revealing consistent fluctuations that generally remain below 80 $\mu g/m^3$. This indicates stable variations in PM2.5 levels without extreme shifts. Figure 8 presents the smoothing effect of the time variable (STOL) on PM2.5 from January 2018 to December 2021, showing an upward trend. This trend suggests a gradual increase in PM2.5 over the study period, indicating a positive association between time progression and PM2.5 levels.

Figure 9 displays the GAMM smoothing effect of COVID-19 new cases on PM2.5, revealing a negative association. As the number of new COVID-19 cases increases, PM2.5 levels tend to decrease, suggesting an inverse relationship between these variables. In contrast, Figure 10 depicts the smoothing effect of COVID-19 deaths on PM2.5, indicating a stable association with a slight upward trend. This result suggests that as COVID-19 deaths increase, there is a modest rise in PM2.5 levels, reflecting a complex interaction between COVID-19 pandemic death rates and air quality. Together, these findings provide insights into the relationships between time, COVID-19 cases, deaths, and PM2.5 during the study period.

## Conclusion and discussion

This paper employed cluster analysis and GAMM to explore the effects of COVID-19-related factors on PM2.5 levels in Thailand, Singapore, and Indonesia. Cluster analysis grouped these countries based on socioeconomic indices and data availability, with Cluster Three, comprising Thailand, Singapore, and Indonesia, selected for further analysis due to comprehensive COVID-19 and PM2.5. GAMM results for Thailand showed no significant association between PM2.5 and COVID-19 factors, such as new cases, deaths, and vaccinations. Instead, seasonal variations, primarily driven by agricultural burning and industrial activities, were identified as the main contributors to PM2.5 fluctuations. While temporary reductions in PM2.5 were observed during lockdowns, these changes were not sustained, suggesting that COVID-19-related restrictions had limited influence on air quality in Thailand.

In contrast, the GAMM results for Singapore and Indonesia revealed significant associations between PM2.5 and COVID-19-related factors, including vaccinations, the stringency index, and case counts. Specifically, the data suggested that as vaccination rates increased and mobility restrictions tightened, PM2.5 levels tended to decrease. This pattern reflects a connection between mobility and economic activity due to pandemic-related measures and the observed improvements in air quality. PM2.5 levels in these countries were also responsive to changes in COVID-19 cases and deaths, implying that public health measures may have contributed to the temporary reduction in pollution. These findings highlight a divergence in the influence of COVID-19 measures on PM2.5 across the three countries studied, with Singapore and Indonesia showing greater sensitivity to health-related restrictions compared to Thailand.

The specific environmental and policy contexts in each country may help explain the observed differences in PM2.5 responses. In Thailand, PM2.5 is primarily driven by seasonal factors such as agricultural burning and industrial emissions, which may have outweighed the impact of COVID-19 restrictions. This aligns with studies, such as Wetchayont (2021), which emphasize the role of seasonal emissions in shaping Thailand's air quality. Conversely, the significant impact of COVID-19 measures on PM2.5 in Singapore and Indonesia is consistent with findings from Barouki et al. (2021) and Gkatzelis et al. (2021), who reported that reduced mobility and economic activity during lockdowns contributed to lower pollution levels in urban areas. These results suggest that public health measures can indirectly affect air quality, their impact varies depending on local pollution sources and policy responses.

These findings have important implications for air quality management policies during health crises. In Singapore and Indonesia, where COVID-19 measures influenced PM2.5 levels, policymakers could consider integrating air quality goals into public health response plans. During future health emergencies, implementing temporary restrictions on industrial and vehicular emissions could help control pollution levels, benefiting both public health and environment. In Thailand, where PM2.5 is primarily driven by seasonal pollution, the focus should remain on specific sources, such as agricultural burning, as highlighted by researchers like Kaewrat and Janta (2021). This approach would allow for more effective, region-specific air quality management tailored to local contexts.

This paper demonstrates that the association between COVID-19-related factors and PM2.5 varies across countries due to differences in environmental conditions and policy responses. In Thailand, PM2.5 is mainly influenced by seasonal factors, while in Singapore and Indonesia, COVID-19-related restrictions and health policies impact air quality. These findings underscore the need for air quality strategies that are of tailored to local conditions and pollution sources, as suggested by researchers such as Wong et al. (2024). For effective environmental policy, governments should consider regional characteristics to simultaneously address both public health and air quality goals during health crises.

### *Limitations and Further Research*

This paper has two main constrains due to its use of data mining techniques. First, the data is based on secondary sources, which can result in missing or incomplete datasets. This limitation may affect the accuracy of the statistical methods used. Second, since the paper employs data mining, it does not begin with hypotheses. Instead, the focus is on exploring the associations among variables and understanding the underlying reasons for the observed results.

This paper offers two recommendations for future research. First, future studies could expand the analysis to countries in other regions, such as the European Union and North America. Second, while this paper focuses on data mining for analysis, future studies could incorporate a broader range of secondary data, including more than 36 months of time-series data, to strengthen the findings.

## Reference

Asian Development Bank. (2020). *Basic Statistics 2020*. Retrieved March 18, 2021 from https://www.adb.org/publications/basic-statistics-2020

Augustin, N. H., Musio, M., von Wilpert, K., Kublin, E., Wood, S. N. & Schumacher, M. (2009). Modeling Spatiotemporal Forest Health Monitoring Data. *Journal of the American Statistical Association*, 104(487), pp. 899-911. DOI:10.1198/jasa.2009.ap07058

Barouki, R., Kogevinas, M., Audouze, K., Belesova, K., Bergman, A., Birnbaum, L., Boekhold, S., Denys, S., Desseille, C., Drakvik, E., Frumkin, H., Garric, J., Destoumieux-Garzon, D., Haines, A., Huss, A., Jensen, G., Karakitsios, S., Klanova, J., Koskela, I.-M. & Vineis, P. (2021). The COVID-19 pandemic and global environmental change: Emerging research needs. *Environment International*, 146, 106272. DOI:10.1016/j.envint.2020.106272

Berkeley Earth. (2020). *Regional Average particulate Air Pollution (PM2.5)* [PM 2.5]. http://berkeleyearth.org/air-quality/local/Indonesia/

Brömssen, C. v. (2016). *Workshop in Generalized Additive Models (GAM)*. https://www.slu.se/globalassets/ew/org/centrb/statisticsslu/workshops/2016/an-introduction-to-gam.pptx

Chen, C. (2000). Generalized additive mixed models. *Communications in Statistics - Theory and Methods*, 29(5-6), pp. 1257-1271. DOI:10.1080/03610920008832543

Constantinescu, C. (2019, April 25). Using generalised additive mixed models (gamms) to predict visitors to edinburgh and craigmillar castles. *Technical blog from our data science team*. https://thedatalab.com/tech-blog/using-generalised-additive-mixed-models-gamms-to-predict-visitors-to-edinburgh-and-craigmillar-castles/

Fahrmeir, L. & Lang, S. (2001). Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 50(2), pp. 201-220. http://www.jstor.org/stable/2680887

Frączek, K., Bulski, K. & Chmiel, M. (2023). Assessment of exposure to fungal aerosol in the lecture rooms of schools in the Lesser Poland region. *Archives of Environmental Protection*, *vol. 49*(No 4), 95-102. DOI:10.24425/aep.2023.148688

Gkatzelis, G. I., Gilman, J. B., Brown, S. S., Eskes, H., Gomes, A. R., Lange, A. C., McDonald, B. C., Peischl, J., Petzold, A., Thompson, C. R. & Kiendler-Scharr, A. (2021). The global impacts of COVID-19 lockdowns on urban air pollution: A critical review and recommendations. *Elementa: Science of the Anthropocene*, 9(1). DOI:10.1525/elementa.2021.00176

Godłowska, J., Kaszowski, K. & Kaszowski, W. (2022). Application of the FAPPS system based on the CALPUFF model in short-term air pollution forecasting in Krakow and Lesser Poland. *Archives of Environmental Protection*, 48(3), pp. 109-117. DOI:10.24425/aep.2022.142695

Groll, A. & Tutz, G. (2012). Regularization for generalized additive mixed models by likelihood-based boosting. *Methods Inf Med*, 51(2), pp. 168-177. DOI:10.3414/me11-02-0021

Hastie, T. & Tibshirani, R. (2017). *Generalized additive models*. Routledge.

Kaewrat, J. & Janta, R. (2021). Effect of COVID-19 Prevention Measures on Air Quality in Thailand. *Aerosol and Air Quality Research*, 21(3), 200344. DOI:10.4209/aaqr.2020.06.0344

Kotsiou, O. S., Kotsios, V. S., Lampropoulos, I., Zidros, T., Zarogiannis, S. G. & Gourgoulianis, K. I. (2021). PM2.5 Pollution Strongly Predicted COVID-19 Incidence in Four High-Polluted Urbanized Italian Cities during the Pre-Lockdown and Lockdown Periods. *International Journal of Environmental Research and Public Health*, 18(10), 5088. DOI:10.3390/ijerph18105088

Lee, M. & Finerman, R. (2021). COVID-19, commuting flows, and air quality. *Journal of Asian Economics*, 77, 101374. DOI:DOI:10.1016/j.asieco.2021.101374

Li, J., Hallsworth, A. G. & Coca-Stefaniak, J. A. (2020). Changing Grocery Shopping Behaviours Among Chinese Consumers At The Outset Of The COVID-19 Outbreak. *Tijdschrift Voor Economische En Sociale Geografie*, 111(3), pp. 574-583. DOI:10.1111/tesg.12420

Liao, Q., Yuan, J., Dong, M., Yang, L., Fielding, R. & Lam, W. W. T. (2020). Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. *J Med Internet Res*, 22(5), e18796. DOI:10.2196/18796

Lim, Y. K., Kweon, O. J., Kim, H. R., Kim, T.-H. & Lee, M.-K. (2021). The impact of environmental variables on the spread of COVID-19 in the Republic of Korea. *Scientific Reports*, 11(1), 5977. DOI:10.1038/s41598-021-85493-y

Liu, Q., Xu, S. & Lu, X. (2021). Association between air pollution and COVID-19 infection: evidence from data at national and municipal levels. *Environ Sci Pollut Res Int*, 28(28), pp. 37231-37243. DOI:10.1007/s11356-021-13319-5

Lorenzo, J. S. L., Tam, W. W. S. & Seow, W. J. (2021). Association between air quality, meteorological factors and COVID-19 infection case numbers. *Environmental Research*, 197, 111024. DOI:10.1016/j.envres.2021.111024

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C. & Rodés-Guirao, L. (2021). A global database of COVID-19 vaccinations. *Nature Human Behaviour*, 5(7), pp. 947-953. DOI:10.1038/s41562-021-01122-8

Meo, S. A., Abukhalaf, A. A., Alessa, O. M., Alarifi, A. S., Sami, W. & Klonoff, D. C. (2021). Effect of Environmental Pollutants PM2.5, CO, NO2, and O3 on the Incidence and Mortality of SARS-CoV-2 Infection in Five Regions of the USA. *International Journal of Environmental Research and Public Health*, 18(15), 7810. DOI:10.3390/ijerph18157810

The World Air Quality Index Project. (2022). *Air Quality Historical Data Platform*. https://aqicn.org/data-platform/register

Tuerlinckx, F., Rijmen, F., Verbeke, G. & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. 59(2), pp. 225-255. DOI:10.1348/000711005X79857

Valdés Salgado, M., Smith, P., Opazo, M. A. & Huneeus, N. (2021). Long-Term Exposure to Fine and Coarse Particulate Matter and COVID-19 Incidence and Mortality Rate in Chile during 2020. *International Journal of Environmental Research and Public Health*, 18(14), 7409. https://www.mdpi.com/1660-4601/18/14/7409

Wang, J., Wang, J. X. & Yang, G. S. (2020). The Psychological Impact of COVID-19 on Chinese Individuals. *Yonsei Med J*, 61(5), pp. 438-440. DOI:10.3349/ymj.2020.61.5.438

Webster, K. (2020). *How COVID 19 changed consumers' daily lives*. Retrieved August 4, 2021 from https://www.pymnts.com/coronavirus/2020/consumer-spending-behavior-covid19-study/

Weisberg, H. F. (1992). *Central tendency and variability*. Sage Publications.

Wetchayont, P. (2021). Investigation on the Impacts of COVID-19 Lockdown and Influencing Factors on Air Quality in Greater Bangkok, Thailand. *Advances in Meteorology*, 6697707. DOI:10.1155/2021/6697707

Wong, W. M., Tzeng, S.-Y., Mo, H.-F. & Su, W. (2024). Predicting air quality trends in Malaysia's largest cities: the role of urban population dynamics and COVID-19 effects. *Archives of Environmental Protection*, 50(2), pp. 65-74. DOI:10.24425/aep.2024.150553

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4), pp. 1025-1036. DOI:10.1111/j.1541-0420.2006.00574.x

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1), pp. 3-36. DOI:10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models: An introduction with R* (2nd ed.). Taylor & Francis. (DOI:10.1201/9781315370279)