PAPER • OPEN ACCESS

A Time-Series Analysis on the Covid-19 Mortality, PM2.5 Levels, and Weather Variables in Denpasar City, Indonesia

To cite this article: Amari Dihaning Suhendarto *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1098** 012020

A Time-Series Analysis on the Covid-19 Mortality, PM2.5 Levels, and Weather Variables in Denpasar City, Indonesia

Amari Dihaning Suhendarto[1], Yeny Dhokhikah[1*], and Abdur Rohman[1]

[1]Study Program of Environmental Engineering, Dept. of Civil Engineering Faculty of Engineering, Universitas Jember
Jalan Kalimantan 37, Jember, Jawa Timur, Indonesia 68121

yeny.teknik@unej.ac.id

Abstract. The objective of this study is to explore the relationship between Covid-19 mortality cases and environmental variables, namely PM2.5 concentration and weather variables, in Denpasar City, Indonesia. Regression models were used. The response variable was the monthly Covid-19 mortality from March 2020 to December 2021 and the predictor variables were the mean concentration of PM2.5, temperature, wind speed, rainfall and duration of sunshine. All data analyzed were provided by the Indonesian Government. Simple linear regression (SLR) and dynamic regression with ARIMA error models were used. Further, of the 22 monthly data, the first 19 months data were used to train the models and the remaining data were used as the test data. It is found that both wind speed and the interaction between PM2.5 concentration and wind speed have statistically significant relationships with Covid-19 mortality. The estimates of SLR and ARIMA (0,1,1) with interaction models show that on average, in case of 0.5 m/s wind speed, an increase of 1 μg/m3 in the monthly mean of daily PM2.5 concentrations associates with 17.4 and 16.3 increase in the monthly Covid-19 mortality case, respectively. Although this study is observational, its findings suggest the importance of controlling PM2.5 concentration.

1. Introduction

The Covid-19 pandemic was first declared by the WHO (World Health Organisation) on 12 March 2020 after the death cases from Covid-19 increased significantly. This pandemic started in Wuhan, Hubei Province of China on 12th December 2019, which caused 80 deaths of 2,794 laboratory-confirmed infections by 26 January 2020 [1][2]. Covid-19 is officially named by WHO to describe the disease caused by the new coronavirus, whilst the International Committee for Taxonomy of Viruses (ICTV) established that the official name of the disease is Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Our World in Data* reported that the cumulative confirmed Covid-19 cases were found in almost all countries, with the five highest cumulative cases per 30 December 2021 were recorded in the United States, India, United Kingdom, France, and Germany [3]. Infected persons might have different responses to Covid-19. Most people who are exposed to this virus will experience mild to moderate symptoms and will recover without needing to be hospitalized. However, some people will experience severe pain and require medical assistance. Covid-19 can spread directly or indirectly, such as through respiratory droplets of an infected person when she speaks, sings, breathes, coughs, or sneezes. The droplets can be coarse aerosols with aerodynamic diameter greater than 5 µm or fine-particle aerosols with aerodynamic diameter less than 5 µm [4]. Some climate conditions (weather, air pollution, temperature, and humidity level) contribute to the increase in confirmed cases of Covid-19 and total death in Milan, Italy [5]. A similar study conducted in Turkey analyzed the correlation between weather (temperature, dew point, humidity, and wind speed), population, and the number cases of Covid-19. The authors stated that high wind speed and low temperature correlate with the number of Covid-19 cases [5]. To examine the correlation between weather and the number of cases or deaths of Covid-19, some models are applied, namely Spearman's correlation analysis [6][7], multiple regression model [8], Pearson's correlation coefficient, linear and non-linear regression, LOESS, two-way ANOVA, and others [9].

This study aims to discover the correlation between the deaths from Covid-19, the concentration of PM2.5, and the weather variables in Denpasar, Indonesia using simple linear regression and ARIMA models.

2. Methods

The area of study in this research is Denpasar, the capital city of Bali province, Indonesia. Denpasar is geographically located between 8°35'31"S to 8°44'49"S and 115°00'23"E to 115°16'27"E, with a total area of 127,78 square kilometers. In 2021 the population of Denpasar was 726,599 and its population density was 5,686 per square kilometer [10].

2.1. Datasets

The datasets of this study were provided by three separate institutions of the Government of Indonesia. The monthly mortality cases data due to Covid-19 were provided by the Public Health Department of Denpasar City. Regarding the predictor variables, the daily concentration of PM2.5 (in micrograms per m3), the daily speed of the wind (in microgram/m^3), the daily wind speed (in m/s), and the daily temperature (^{0}C) were provided by the Bali and Nusa Tenggara Ecoregional Control and Development Directorate, Ministry of Environment and Forestry. The daily precipitation records (in mm) and the daily duration of the sunshine (in hour/day) were provided by BMKG (Meteorological, Climatological, and Geophysical Agency). Because mortality case data were available only in monthly intervals, not daily ones, all daily data of the predictor variables were summarised to obtain a mean value for each month. Missing values were excluded from the summarization step. In total, there were 22 data points for each variable, representing the corresponding 22 months of observations (March 2020 to December 2021). The datasets were then divided into two groups, namely, training group and testing group. The training group consisted of data points from March 2020 to September 2021 and was used to build the models. The testing group consisted of data points from October to December 2021 and was used to measure and compare the accuracy of the models built on the training data. The summary of the 22-monthly datasets is presented in Table 1.

Table 1. The data summaries of the variables used in this study

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
Mortality case (people)	0.00	4.25	21.00	44.91	52.00	357.00	78.76
PM2.5 concentration (μg/m^3)	7.73	9.91	12.16	12.33	14.20	18.70	3.13
Air Temperature (°C)	26.21	26.87	27.14	27.19	27.56	28.32	0.58
Wind speed (m/s)	0.30	0.45	0.67	0.82	1.07	2.04	0.47
Rainfall (mm)	0.03	2.87	6.45	8.20	10.90	28.56	7.93
Sunshine duration (hour/day)	3.31	5.71	7.03	6.50	7.40	8.41	1.39

For the mortality case, the mean value was 45 cases/month, but the number of cases varied widely, as indicated by the large standard deviation. The highest mortality case occurred in August 2021 with 357 cases. The significant rise occurred shortly after the outbreak of the *delta* variant of the Covid-19 virus in July 2021. The mean daily concentration of PM2.5 varied in the range of 8 to 19 μg/m^3. The mean daily temperature over the months was relatively stable in the range of 26 to 28 degree Celsius. The mean daily wind speed was low, mostly lower than 2 m/s. A closely similar time series pattern was found between the wind speed data and the mortality case from Covid-19. Because Indonesia has a dry season and a rainy season, the rainfall varied as expected with a mean daily value of 8.20 mm and the highest values occurred during the usual rainy season period, namely around November to January each year. The duration of the daily sunshine varied slightly, with a mean value of 6.5 hours/day and a standard deviation of 1.4 hours/day. This value was also expected because Denpasar is located relatively close to the Equator.

2.2 Simple Linear Regression

In a simple linear regression, the response variable is modelled as a linear function of the predictor variables with an error term assumed to be independent and identically distributed random variables that follow a normal distribution with constant variance, as presented in equation (1)

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad (1)$$

where y denotes the response variable, x is the response variable, β denotes the coefficient of the predictor variables. The values of the coefficients are estimated from the observation data using the least-squares estimation. Each coefficient represents the change in the mean response variable for every unit increase in the corresponding predictor variable when all the other predictors are kept constant. The independence of the errors is evaluated in this study by testing the serial correlation of the residuals of the model because the presence of serial correlation implies that the residuals, and hence the errors, are not independently distributed. The variables in the regression are assumed to be independent of each other. The response variable in the Simple Linear Regression is the monthly mortality case due to Covid-19, and the predictor variables are the other five variables presented in Table 1.

2.3. Dynamic Regression Model with Autoregressive Integrated Moving Average (ARIMA)

An ARIMA model is constructed from the combination of differencing, autoregression, and a moving average model. Integration is the inverse of differencing. The entire model can be written as in equation (2)

$$y_{t,d} = \mu + \alpha_1 x_{1,d} + \alpha_2 x_{2,d} + \ldots + \alpha_n x_{n,d} + \beta_1 y_{t-1} + \beta_2 y_{t-2,d} + \ldots + \beta_p y_{t-p,d}$$

$$+\epsilon_t + \lambda_1 \epsilon_{t-1} + \lambda_2 \epsilon_{t-2} + \ldots + \lambda_q \epsilon_{t-q} \quad (2)$$

where $y_{t,d,p}$ denotes the response variable at timestamp *t*, lagged with *p* order, and differenced with *d* order, μ denotes the intercept, $x_{i,d}$ denotes the predictor variable *i* that is differenced with *d* order, and α, β, and λ denote coefficients of the predictor variables. Their values are estimated from the observation data by using least squares estimation, ϵ_{t-q} denotes the error lagged with q order. This is known as an ARIMA(p,d,q) model.

3. **Results and Discussion**

3.1. Serial Correlation in Each Variable Data

The presence of serial correlation, or, in another term, autocorrelation, in each variable data was inspected through the ACF (Autocorrelation Function) plot and PACF (Partial Autocorrelation) plots. It is observed that a significant serial correlation was found in each variable. In addition, in all variables the lag = 1 shows significant serial correlation, indicating that ARIMA models with d = 1 differencing are highly probably appropriate to represent the serially correlated data.

3.2. Goodness-of-Fit of the Models

Based on the aforementioned ACF and PACF plots, we examined ARIMA (0,1,0) and several models with relatively close specifications. The AIC (Akaike Information Criterion) and the corrected AIC (AICc) were used to assess the goodness of fit of each model with the observation data. The corrected AIC is also used in this study because the sample size (n = 22) can be considered as a small sample size, in which AICc is recommended because it takes the number of parameters into consideration [11]. Table 2 shows that ARIMA (0,1,1) is the model with the best-fit based on AIC values. However, based on the corrected AIC, ARIMA (0,1,0) is the best model. Both ARIMA models are better suited than the SLR model. Therefore, we further analyse the three models, namely the SLR, ARIMA (0,1,0) and ARIMA(0,1,1).

Table 2. Estimates, Standard Errors, and p-values in the SLR, ARIMA (0,1,0), and ARIMA(0,1,1) models without interaction.

Term	Simple Linear Regression			ARIMA (0,1,0)			ARIMA (0,1,1)		
	estimate	std. error	p.val	estimate	std. error	p.val	estimate	std. error	p.val
Intercept	775.48	844.72	0.375	-5.03	13.65	0.717	0.58	3.01	0.849
MA(1)	undefined			undefined			-1.00	0.17	0.000
Temperature	-32.08	32.12	0.336	3.74	37.02	0.921	-30.61	28.33	0.294
PM2.5	9.56	6.39	0.158	1.25	7.80	0.875	9.21	5.73	0.125
WindSpeed	181.21	40.89	0.000	231.97	50.37	0.000	175.35	46.09	0.001
Rainfall	2.10	3.73	0.582	–0.24	2.99	0.935	1.90	3.33	0.575
SunshineDur	-21.46	22.06	0.348	-32.09	22.47	0.170	-21.15	18.80	0.275
AIC	213.11			210.27			208.52		
Corrected AIC	223.30			221.47			224.52		

Our analyses consist of two steps. First, we analysed the three models without any interaction between the predictor variables. The results are shown in Table 3. It is found that in the three models, the mean daily wind speed is the only variable that is statistically highly significant (with p-values far lower than 0.05). It is not reasonable that wind speed alone is associated with mortality due to Covid-19 because the disease is caused by a virus. Based on [11], we suspect that the concentration of PM2.5 probably interacts with the wind speed during virus transmission.

Therefore, in the second step, we analysed the models including the interaction term between PM2.5 concentration and wind speed. The results show that all models with interaction terms are better suited than those without an interaction term, as seen in Table 3. Based on the AIC values, when the interaction term is included, the best-fit model is ARIMA(0,1,1), but based on the corrected AIC values, the SLR model is the best-fit model.

Table 3. Estimates, Standard Errors, and P-values in the SLR, ARIMA (0,1,0), and ARIMA(0,1,1) with Interaction Terms

Term	Simple Linear Regression			ARIMA (0,1,0)			ARIMA (0,1,1)		
	estimate	std. error	p.val	estimate	std. error	p.val	estimate	std. error	p.val
Intercept	1068.39	675.37	0.139	-1.88	10.79	0.864	4.11	2.25	0.085
MA(1)	undefined			undefined			-1.00	0.14	0.000
Temperature	-54.48	26.51	0.062	-25.54	30.46	0.413	-48.02	20.24	0.028
PM2.5	32.24	9.17	0.004	29.97	10.61	0.011	33.77	6.93	0.000
WindSpeed	502.42	113.11	0.001	580.39	112.19	0.000	518.27	85.30	0.000
Rainfall	3.63	2.99	0.248	1.80	2.44	0.469	2.50	0.05	0.298
SunshineDur	-14.73	17.60	0.418	-4.74	19.52	0.811	-11.31	13.33	0.407
PM2.5 : WindSpeed	-29.68	10.01	0.011	-36.45	10.98	0.003	-34.98	8.06	0.000
Goodness-of-Fit Metrics for the Models									
AIC	204.68			203.67			197.62		
Corrected AIC	219.08			219.67			220.12		

Table 3 shows that in the three models, each concentration of PM2.5, the speed of the wind and the interaction of the two variables have a statistically significant linear relationship with the mortality case due to Covid-19 with slightly different slopes between the models. Due to the presence of the interaction, to interpret the relationship, the coefficients of the interaction term must be accounted for. For example, the SLR models implies that on average, in case of 0.5 m/s wind speed, and other variables are held constant, an increase of 1 μg/m3 in mean daily PM2.5 concentration in a month associates with 32.24 - (0.5) (29.68) = 17.4 increase in mortality case due to Covid-19 per month. As for ARIMA (0,1,0) and ARIMA (0,1,1), the models implies association between 1 μg/m3 in mean daily PM2.5 concentration in a month with an 11.7 and 16.3 increase in mortality case due to Covid-19 per month. Likewise, the SLR implies that, in case of PM 2.5 concentration is 10 μg/m3 and other variables are held constant, an increase in daily mean wind speed of 1 m/s associates with 502.42 - (10) (32.24) = 180 increase in mortality case due to Covid-19 per month, on average.

The temperature is found to be statistically significant in the ARIMA (0, 1,1) model, but not statistically significant in other models. In all models, the coefficient of the temperature variable is negative, which implies that the increase in temperature is associated with a decrease in mortality due to Covid-19. The rainfall variable and the duration of sunshine are not statistically significant in all models. The fitness of the models in each month data are visually presented in Figure 1.

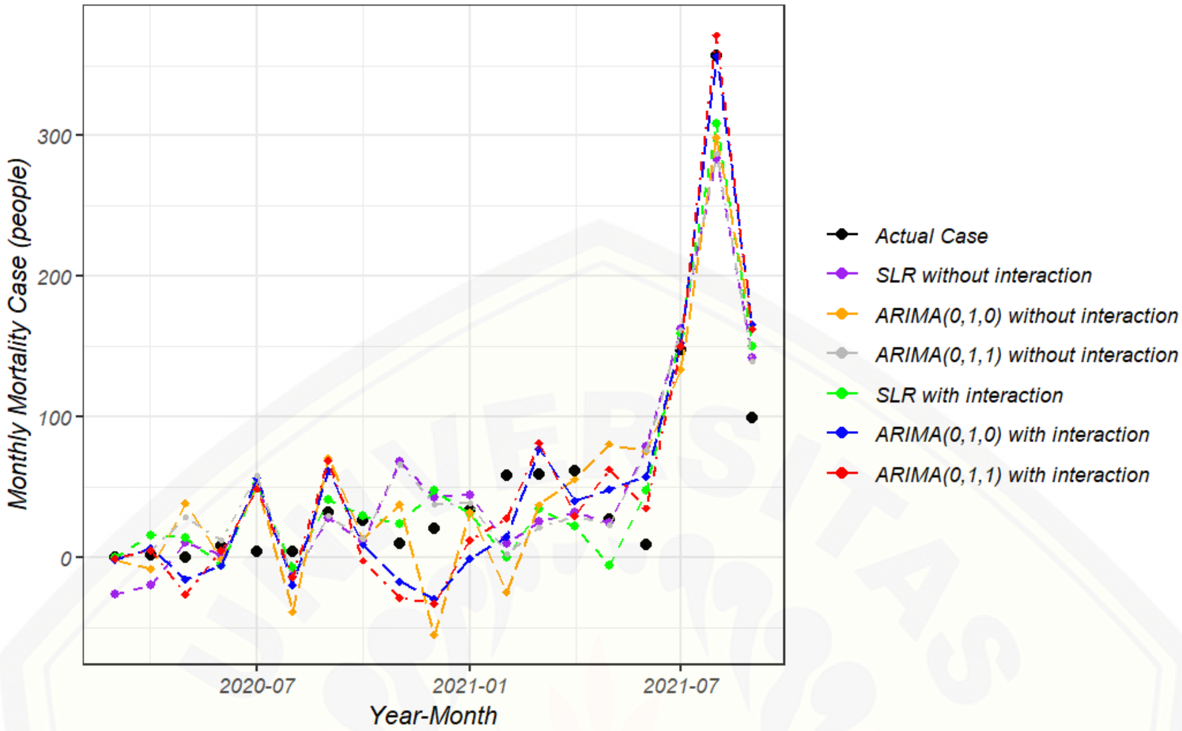

Figure 1. Actual values of monthly mortality cases and the corresponding fitted values of the models

3.3. Accuracy of the Models

After the goodness-of-fits of the six models are evaluated based on their AIC and the corrected AIC values, the accuracies of the models are compared. The comparison is conducted by applying each model to the test data of the predictor variables to generate the predicted mortality case for the last three months of 2021. The accuracy is measured by using the Mean Absolute Error (MAE). The comparison results are presented in Table 4.

Table 4. Predicted mortality cases due to Covid-19 of SLR, ARIMA (0,1,0), and ARIMA (0,1,1) models

		without interaction			with interaction		
Month	Actual Case	SLR	ARIMA (0,1,0)	ARIMA (0,1,1)	SLR	ARIMA (0,1,0)	ARIMA (0,1,1)
2021 Oct	21	269	157	266	210	108	58
2021 Nov	5	49	-86	41	56	-106	-154
2021 Dec	2	79	-30	70	46	-86	-126
Mean Absolute Error		123.0	86.3	116.3	94.7	95.3	108.0

Table 4 shows that each model results in a much larger mean absolute error than the actual case. This poor accuracy can be attributed to the small sample size of the model and the highly uncertain nature of Covid-19 development. Table 4 also shows that the best-fit model in the training phase does not always show the best accuracy in the testing phase.

3.4. Comparison with previous studies
This study found that each of the concentration of PM2.5, the wind speed, and the interaction of the two variables have a significant relationship with the case of mortality from Covid-19. The relationship between particulate matter concentration and Covid-19 has been strongly supported by the finding in [13] in which SARS-CoV-2 RNA was found in PM10 samples in Bergamo, Italy. In another study [14], it was found that other things being equal, a municipality in the Netherlands with 1 μg/m^3 more PM2.5 concentrations will have 9.4 more Covid-19 cases, 3.0 more hospital admissions, and 2.3 more deaths. These findings are clearly in agreement with the findings of our study.

Since our study is observational, no causal relationship can be inferred from the results. Nevertheless, our findings are in close agreement with the findings in [15] in which it is found that mean wind speed has a significant correlation with Covid-19 confirmed cases in Bandung City, Indonesia. However, in [15] it is found that the mean temperature has a significant correlation with the Covid-19 confirmed case, while in our study, the relationship between the mean temperature and the Covid-19 mortality case is not significant in most models.

4. **Conclusion**
The Covid-19 pandemic spread throughout the world in a few months since it exploded in Wuhan Province of China in late 2019. This study analysed six models consisting of Simple Linear Regression (SLR) and ARIMA models. Weather conditions and air pollution are found to have contributed to the increasing case of the Covid-19 mortality rate. Each of the PM2.5 concentration, the wind speed and the interaction between the PM2.5 concentration and the wind speed have a statistically significant relationship with the Covid-19 mortality cases. Temperature is not statistically significant in most models. The SLR, ARIMA (0,1,0) and ARIMA(0,1,1) with interaction models implies that on average, in case of 0.5 m/s wind speed, and other variables are held constant, an increase of 1 μg/m3 in mean daily PM2.5 concentration in a month associates with 17.4, 11.7, and 16.3 increase in mortality case due to Covid-19 per month, respectively. Our study is observational, therefore, the relationships found between Covid-19 mortality case and PM2.5 concentration and wind speed cannot be claimed to be causal. However, our findings are in agreement with previous studies in Italy, the Netherlands, and Indonesia. These findings suggest the importance of PM2.5 concentration control to hinder the spread of Covid-19.

References
[1] Zhou P, Yang XL, Wang XG, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579 (7798):270–273. https://www.nature.com/articles/s41586-020-2012-7
[2] Ciotti M, Ciccozzi M, Terrinoni A, Jiang WC, Wang CB, Bernardini S. 2020. The Covid-19 pandemic. Critical Reviews in Clinical Laboratory Sciences. 57 (6): 365-388. https://www.tandfonline.com/doi/full/10.1080/10408363.2020.1783198
[3] Our World in Data. 2022. *Cumulative confirmed COVID-19 cases, Dec 30, 2021*. https://ourworldindata.org/coronavirus#coronavirus-country-profiles (Accessed: 18 September 2022).
[4] Leung NHL, Chu DKW, Shiu EYC, Chan KH, McDevitt JJ, Haul BJP, Yen HL, Li Y, Ip DKM, Peiris JSM, Setol WH, Leung GM, Milton DK and Cowling BJ. 2020. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nature Medicine. 25 (5): 676-680. https://www.nature.com/articles/s41591-020-0843-2
[5] Zoran MA, Savastru RS, Savastru DM, Tautan MN. 2020. Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. Science Total Environment. 738: 1-12. https://www.sciencedirect.com/science/article/pii/S0048969720333453
[6] Sahin M. 2020. Impact of weather on COVID-19 pandemic in Turkey. Science of The Total Environment. 728 (8): 1-5. https://www.sciencedirect.com/science/article/pii/S0048969720323275

[7] Tosepu R, Gunawan J, Effendy SD, Ahmad LOAI, Lestari H, Bahari H, Asfian P. 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Science of The Total Environment. 725: 1-4. https://www.sciencedirect.com/science/article/pii/S0048969720319495

[8] Rath S, Tripathy A, Tripathy AR. 2020. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 14 (5): 1467-1474. https://www.sciencedirect.com/science/article/pii/S1871402120302939

[9] Ganegoda NC, Wijaya KP, Amadi M, Erandi KKWH & Aldila D. 2021. Interrelationship between daily COVID-19 cases and average temperature as well as relative humidity in Germany. Scientific Report. 11:11302. https://www.nature.com/articles/s41598-021-90873-5

[10] BPS-Statistics of Denpasar Municipality. 2022. Denpasar Municipality in Figures 2022. https://denpasarkota.bps.go.id/publication/2022/02/25/68f4c38625094b798b0471a6/kota-denpasar-dalam-angka-2022.html (Accessed: 18 September 2022).

[11] Dyer, O. 2021. Covid-19: Indonesia becomes Asia's new pandemic epicentre as delta variant spreads. BMJ (Clinical Research Ed). 16 July 2021. Available at https://www.bmj.com/content/bmj/374/bmj.n1815.full.pdf . Accessed on August 8, 2022.

[12] Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. https://otexts.com/fpp3 . Accessed on 20 April 2022.

[13] Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Perrone, M. G., Borelli, M., Palmisani, J., Di Gilio, A., Torboli, V., Fontana, F., Clemente, L., Pallavicini, A., Ruscio, M., Piscitelli, P., & Miani, A. (2020). SARS-Cov-2RNA found on particulate matter of Bergamo in Northern Italy: First evidence. *Environmental Research*, *188*(May), 109754. https://doi.org/10.1016/j.envres.2020.109754

[14] Cole, M. A., Ozgen, C., & Strobl, E. (2020). Air Pollution Exposure and Covid-19 in Dutch Municipalities. *Environmental and Resource Economics*, *76*(4), 581–610. https://doi.org/10.1007/s10640-020-00491-4

[15] Azmy, A.R and Wulandari, R.A. 2022. Hubungan Cuaca dengan Kasus Terkonfirmasi Aktif COVID-19. Media Kesehatan. Available at https://journal.poltekkes-mks.ac.id/ojs2/index.php/mediakesehatan/article/view/2459