Prediction and Analysis of The Number of Ari Cases Based on $Pm_{2.5}$ Concentration With Co-Kriging Approach

Nur Chamidah[1*], Putu Eka Andriani[2], Marfa Audilla Fitri[3], Sofia Andika Nur Fajrina[4], Alda Fuadiyah Suryono[5], Victoria Anggia Alexandra[6]
[1,2,3,4,5,6]Statistics Study Program, Department of Mathematics, Universitas Airlangga, Surabaya, Indonesia
*Corresponding author: nur-c@fst.unair.ac.id

Received: 31 May 2024
Revised: 20 June 2024
Accepted: 26 June 2024

ABSTRACT – Air quality significantly impacts global environmental health, influencing both human well-being and climate change. According to the World Health Organization (WHO), air pollution is one of the most substantial environmental threats to human health, with Indonesia experiencing particularly severe air quality issues. The World Air Quality Report ranks Indonesia 14th globally and 1st in Southeast Asia for poor air quality, with a notable increase in $PM_{2.5}$ concentrations to 37.1 μg/m³ in 2023. Major sources of pollution include coal-fired power plants, motor vehicles, forest fires, and agricultural activities. In urban areas like Surabaya, $PM_{2.5}$ levels have risen, contributing to high incidences of Acute Respiratory Infections (ARI). Spatial analysis reveals a correlation between $PM_{2.5}$ levels and ARI cases, with spatial regression and co-kriging methods offering accurate estimation models. This study utilizes co-kriging, incorporating $PM_{2.5}$ data from nine districts in Surabaya, to estimate ARI cases. The Exponential semivariogram model provided the most accurate predictions, with a MAPE value of 5.11%. The highest estimated ARI cases were in the Kenjeran district, highlighting the need for targeted interventions. Future research should expand observation points and consider additional influencing factors such as weather, population density, and socioeconomic conditions to enhance prediction accuracy and support effective public health strategies.

Keywords– **Co-Kriging, Spatial Data, Acute Respiratory Infection, $PM_{2.5}$**

I. INTRODUCTION

Air quality is a global environmental issue that significantly affects the quality of life for living beings. According to the World Health Organization (WHO) [1],air pollution is one of the largest environmental threats to human health. Also, poor air quality directly impacts the environment, such as causing climate change. Based on the World Air Quality Report [2], Indonesia ranks 14th globally and 1st in Southeast Asia for the worst air quality. Indonesia's average annual $PM_{2.5}$ concentration sharply increased in 2023 to 37.1 μg/m³, a rise of more than 20% compared to 2022. According to Law No. 23 of 2007 on Environmental Pollution, air pollution is caused by human activities such as emissions from factories, motor vehicles, burning trash, agricultural residues, and natural events like forest fires and volcanic eruptions that release dust, gas, and ash clouds. Most air pollutants in Indonesia come from coal-fired power plants, forest fires, and peatland clearing for agricultural development [2]. Acute air pollution occurs during the dry season, typically from July to September, but it can be influenced by changing meteorological conditions.

Emissions from the combustion of motor vehicle fuels also cause air pollution. Urban areas are the largest emitters of carbon emissions from motor vehicle combustion. One metropolitan city with serious air quality issues is Surabaya. The average $PM_{2.5}$ concentration in Surabaya in 2023 ranked 11th in Indonesia, with an average of 27.6 μg/m³, which falls into the moderate category [2]. The PM2.5 concentration in Surabaya has tended to increase in 2023 and 2024, warranting caution. According to WHO [1], there are 7 million premature deaths each year due to the combined effects of outdoor and household air pollution, with millions more falling ill from breathing polluted air. According to the WHO Global Air Quality Guidelines [2], fine particulate matter ($PM_{2.5}$) can penetrate the lungs and enter the bloodstream, affecting all major organs. Exposure to $PM_{2.5}$ can cause cardiovascular and respiratory diseases, such as stroke, lung cancer, and chronic obstructive pulmonary disease (COPD). Air pollution is the fourth leading cause of premature death worldwide, accounting for 29% of lung cancer deaths, 17% of acute respiratory infection deaths, and 43% of COPD deaths.

The increase in air pollution has had a negative impact on health, especially concerning respiratory diseases. According to Nanik Sukristina, the Head of the Surabaya Health Office [3], cases of Acute Respiratory Infections (ARI) in Surabaya reached 174,222 during the first semester of 2023. This data is cumulative from all health service facilities in Surabaya. This figure shows that ARI cases are very high, having increased significantly compared to the same period in 2022. The implementation of Sustainable Development Goals (SDGs) in addressing diseases caused by air pollution is crucial to achieving global health goals holistically. This relates to SDG Goal 3: Good Health and Well-being and Target 3.9: Reducing illnesses and deaths from hazardous chemicals and pollution. To address this issue, efforts to reduce pollution and promote a clean environment are essential. This aligns with several other SDG targets, such as Goal 6 (Clean Water and Sanitation), Goal 7 (Affordable and Clean Energy), and Goal 13 (Climate Action), as pollution in water, air, and energy are major causes of health and environmental issues.

Observations of ARI cases and $PM_{2.5}$ concentrations in Surabaya indicate a spatial dependency. This shows a correlation between the number of ARI cases and PM2.5 concentration conditions in Surabaya, viewed from spatial characteristics such as patterns, size, distance, and regional conditions. Furthermore, each observation location in Surabaya has its unique characteristics regarding ARI data and $PM_{2.5}$ concentrations, requiring better analysis than linear regression. Spatial regression can be used as a solution for modeling data with spatial elements. Spatial regression is a statistical method developed from classical linear methods due to the influence of location or spatial factors on the analyzed data [4]. Therefore, spatial analysis is a suitable method for estimating the number of ARI cases in Surabaya.

To estimate data at unsampled locations, a technique is needed. One geostatistical technique used for estimation and interpolation of data at unsampled locations is called kriging. Several common kriging methods include ordinary kriging and co-kriging [5]. Estimating unobserved ARI cases in Surabaya can be done using spatial analysis methods. With $PM_{2.5}$ concentration as secondary data in estimating ARI cases, spatial analysis with the co-kriging method is required. Co-kriging is a method that not only relies on spatial correlation but also utilizes secondary control data as a corrector for the primary attribute to be estimated [6]. The co-kriging method involves two variables in performing spatial interpolation [7]. The co-kriging method was chosen in the study [8] because it is considered more accurate than the kriging method.

Previous studies have provided a foundation for further analysis. For example, a study by [8] on the application of kriging and co-kriging in spatial estimation of groundwater quality parameters with Sodium Adsorption Ratio (SAR) as the primary variable and Chloride (C_l) as the secondary variable showed that the co-kriging method is more accurate than the kriging method. However, both methods generally had appropriate accuracy for estimating SAR and C_l based on water salinity parameters. A study by [9] on estimating NO_2 concentration using the co-kriging method in Jakarta with NO_2 content as the primary variable and SO_2 content as the secondary variable in eight Jakarta areas showed that NO_2 content in Tanjung Priok, a coastal area, significantly deviated from estimates in other Jakarta areas, ranking highest among the locations studied, with GBK showing the lowest content with co-kriging estimates. Then, a study by [10] on applying the co-kriging method in predicting monthly rainfall in West Java with rainfall as the primary variable and elevation as the secondary variable also showed that applying the co-kriging method with rainfall as the primary variable and elevation as the secondary variable produced satisfactory results.

Based on explanations from previous studies and the background described, the researcher is interested in estimating the number of respiratory disease cases related to air quality in Surabaya using the co-kriging method. This study is expected to enhance the researcher's knowledge of co-kriging and its application and serve as a reference for future research.

II. LITERATURE REVIEW

A. *Particulate Matter 2.5*

Particulate Matter (PM) is a type of hazardous pollutant that varies in size and can cause increased mortality due to exposure to air pollution. Particulate Matter 2.5 ($PM_{2.5}$), also known as fine particle, is a type of particulate matter that is very small in size and can cause various diseases. When inhaled into the body, it can penetrate into the lower respiratory tract and can pass through the bloodstream [11].

Particulate Matter 2.5 ($PM_{2.5}$) is airborne particles that have a size ≤ 2.5μm that can be inhaled, cannot be filtered in the upper respiratory system, and settles in the respiratory tract until it reaches the lungs. Particulate Matter 2.5 ($PM_{2.5}$) comes from various sources such as the combustion of motor vehicle fuel to forest fires [12]. According to the World Health Organization (WHO), $PM_{2.5}$ can also cause respiratory tract infections (URI), lung cancer, cardiovascular disease, premature death and chronic obstructive pulmonary disease

B. *Acute Respiratory Infection (ARI)*

Acute Respiratory Infection (ARI) is an infection that occurs in a component of the respiratory tract. ARI consists of upper respiratory tract infections or upper respiratory tract infections (URTI) and lower respiratory tract infections or lower respiratory tract infections (LRTI). URTI is associated with infections in or above the larynx [13]. ARI includes rhinitis, pharyngitis, epiglottitis, tonsillitis, and laryngitis. Symptoms of ARI generally include cough, sore throat, runny nose, stuffy nose, headache, mild fever, facial pressure, sneezing and myalgia. The onset of symptoms usually begins one to three days after exposure and lasts for 7–10 days, and can last up to 3 weeks [14]. Meanwhile, LRTI is an infection that occurs below the larynx and includes bronchitis, bronchiolitis and pneumonia.

ARI causes inflammation of the respiratory tract, from the nose to the lungs. This condition can be caused by viral and bacterial infections which spread very easily, such as through the sufferer's droplets. ARI is very easily transmitted and can be experienced by anyone, especially children and the elderly. The cause of ARI according to [15] is caused by viruses such as rhinoviruses, RSV, adenovirus, influenza virus or parainfluenza virus. Apart from viruses, there are several types of bacteria that can also cause ARI, including streptococcus and staphylococcus aureus.

There are several factors that can increase a person's risk of contracting ARI, including children under 5 years old or elderly and someone who has a weak immune system. This is because a weak immune system will find it difficult to fight bacteria and viruses that enter the body, making it susceptible to disease. Apart from that, someone who has a smoking habit or is a passive smoker is also at risk of developing ARI because they are often exposed to direct cigarette smoke which can irritate the respiratory tract [16]. Another cause that can increase a person's exposure to ARI is frequent exposure to air pollution. This is because dangerous substances that enter through the nose as a result of air pollution can settle in the respiratory tract so that they can irritate and cause ARI [17].

C. Spatial Data

Spatial data is data resulting from measurements that include information about the location and measurement of an object. This information is usually presented in the form of the geographical position of the object and its relationship with other objects, using coordinate points and areas. According to [18], spatial data falls into the category of dependent data because it is collected from different spatial locations, indicating a dependency between the data measurements and the location. Spatial data allows us to obtain information about the geographic coordinates of each area, as displayed in the form of maps or satellite images.

The difference between spatial data and other types of data is local (spatial) information and descriptive (attribute) information. Local (spatial) information relates to coordinates such as geographic coordinates (latitude and longitude) and coordinates, including projection and datum information. Descriptive (attribute) or non-spatial information is information related to a specific location, such as population, vegetation types, and so on [19].

D. Experimental Semivariogram

In the co-criging analysis, a variogram is performed to evaluate the correlation between the difference in observed values and the distance between the corresponding observation points [20]. The experimental auto-covariance equation can be defined as follows

$$C_1(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (U_i - \bar{U})(U_{i+h} - \bar{U}) \tag{1}$$

$$C_2(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (V_i - \bar{V})(V_{i+h} - \bar{V}) \tag{2}$$

where,
$C_1(h)$ = *auto-covariance* value of the primary variable (u) at distance h
$C_2(h)$ = *auto-covariance* value of the secondary variable (v) at distance h
U_i = observation value of the primary variable (u) at point i
U_{i+h} = observation value of the primary variable (u) at point $i+h$
V_i = observation value of the secondary variable (v) at point i
V_{i+h} = observation value of the secondary variable (v) at point $i+h$
$N(h)$ = the number of pairs of points that have distance h

The experimental cross-covariance equation can be defined as follows:

$$C_{12}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (U_i - \bar{U})(V_{i+h} - \bar{V}) \tag{3}$$

where,
$C_{12}(h)$ = *cross-covariance* value of the primary variable (u) and secondary variable (v) at distance h

E. Theoritical Semivariogram

The experimental auto and cross covariances generated from the data often have irregular patterns, making them difficult to understand and cannot be used directly. Therefore, structural analysis is required to match the experimental covariance with the theoretical covariance [21].

According to McBratney [22], there are three theoretical covariance models that are often used in co-kriging analysis, namely the spherical, exponential, and gaussian models. The form of the three models can be defined as follows.

1.) Spherical Model

$$C(h) = \begin{cases} P+Q & ; h = 0 \\ (P+Q)\left\{1 - 1{,}5\left(\frac{h}{r}\right) - 0.5\left(\frac{h}{r}\right)^3\right\} & ; 0 \le h \le r \\ 0 & ; h > r \end{cases} \tag{4}$$

2.) Exponential Model

$$C(h) = (P+Q)\left[1 - exp\left(-\frac{h}{r}\right)\right] \tag{5}$$

3.) Gaussian Model

$$C(h) = (P+Q)\left[1 - exp\left(-\frac{h}{r}\right)^2\right] \tag{6}$$

where,
P (nugget effect) = approximation of auto covariance and cross covariance values at distances around zero
Q (sill) = maximum value which achieved by auto covariance and cross covariance
r (range) = distance when the covariance has achieved its maximum value

F. Co-Kriging Method

Co-Kriging is an extension of the kriging method that uses more than one variable. The Co-Kriging method is an estimation method that minimizes the variance of the estimation error by utilizing the cross-correlation between two variables, namely the main variable (primary) and additional variables (secondary) [23]. The primary variable is the main variable used to interpolate while the secondary variable is a supplementary variable as additional data for the primary variable in order to obtain more accurate estimation results. Co-Kriging estimation is a linear combination of the main and auxiliary variable data expressed as follows:

$$\hat{u}_0 = \sum_{i=1}^{n} a_i u_i + \sum_{i=1}^{m} b_j v_j \tag{7}$$

where,

$\hat{u}_0$ = estimated u at location 0
$u_1, u_2, \dots, u_n$ = main variable data at n locations
$v_1, v_2, \dots, v_n$ = auxiliary variable data at m locations
$a_1, a_2, \dots, a_n$ = main variable weights
$b_1, b_2, \dots, b_n$ = main variable weights

III. METHODOLOGY

A. Data Source

The data used in this research is secondary data obtained from the January 2024 Air Quality Report on the Nafas app and the Surabaya One Data website. The data includes air quality information in Surabaya and the number of respiratory system diseases in Surabaya as in December 2023.

B. Research Variables

Research variables are aspects possessed by the research subjects, which can be individuals, objects, or events, gathered from the research subjects to describe the condition or characteristics of each research subject. In this study, there are two variables: primary variables and secondary variables. The variables and their operational definitions used in this research are listed in Table 1 as follows:

Table 1 Research of Variables

No.	Variables	Operational Definition
1	Primary Variable	Number of ARI Cases at 9 Community Health Centers in Surabaya
2	Secondary Variables	$PM_{2.5}$ at 9 Stations in Surabaya
3	Variable Distance	Longitude and latitude coordinates of each region

C. Analysis Steps

The process of data analysis in this study is as follows:

1. Pre-process data before analyzing using Co-Kriging Method.
2. Test the spatial autocorrelation assumption using the Moran I test.
3. Test the correlation assumption using the Pearson Correlation test.
4. Estimate model with the Co-Kriging method which begins with the calculation of the matrix $\boldsymbol{h_{ij}}$.
5. Determine the model of the experimental semovariograms and determine the values of $\boldsymbol{P}$ (nugget), $\boldsymbol{Q}$ (sill), and $\boldsymbol{r}$ (range) on each experimental semivariogram.
6. Determine theoretical semovariograms with three models: spherical, exponential, and gaussian.
7. Predict In-Sample and Out-Sample data on each theoretical semivariograms model.
8. Compare the performance of the most optimal semivariogram model with minimum MAPE.
9. Interpolate using the best semivariogram model of the data.
10. Create a point distribution map and an interpolation map of the number of ARI cases.

IV. RESULTS AND DISCUSSIONS

A. Descriptive Statistics

Descriptive statistics were carried out to determine the general description of the characteristics for the $PM_{2.5}$ concentration variables and the number of ARI diseases in several health centers in Surabaya. The following are the results of descriptive statistics from the research variables of this paper.

Table 2. Charasteristic of Variables

No.	Variables	N	Mean	Maximum	Minimum	St.Dev
1	ARI Cases in Health Center	9	188.44	421	64	111.224
2	$PM_{2.5}$ contentration	9	36.89	39	30	3.018

Based on Table 2, it is found that the characteristics of the $PM_{2.5}$ concentration variable in Surabaya have an average of 36.89 with a standard deviation of 3.018. Meanwhile, the average number of ARI cases observed at community health centers in Surabaya was 188.44 with a standard deviation of 111.224.

B. *Point Distribution Map*

Figure 1 Point Distribution Map between PM2.5 Station and Community Health Center In-Sample

Based on Figure 1, a distribution map can be obtained between the $PM_{2.5}$ Station and the observed health centers. The in-sample health centers used were health centers located in Krembangan, Rungkut, Bulak, Sambikerep, Gubeng, Sawahan, Tenggilis Mejoyo, Pakal and Wonocolo sub-districts. Meanwhile, the $PM_{2.5}$ stations from the nine sub-districts used include South Krembangan, Medokan Ayu, Kenjeran, Lontar, Kertajaya, Kedungdoro, Tenggilis Mejoyo, Babat and Jemur Wonosari.

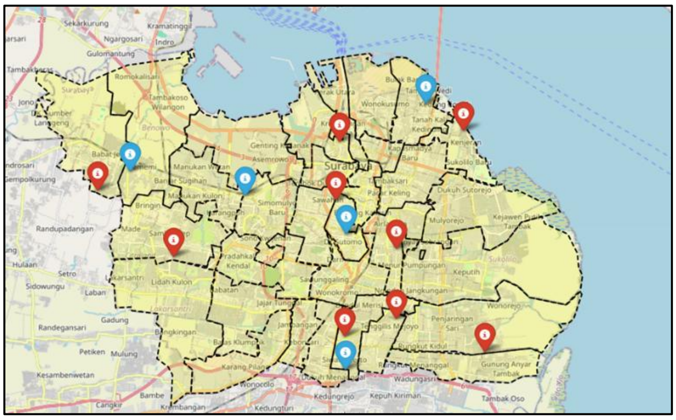

Figure 2 Point Distribution Map between Health Center In-Sample data Health Center and Out-Sample data Health Center

Based on Figure 2, a map of the distribution of locations between the observed health centers and the estimated health centers is obtained. The observed health centers or health centers used for the In-Sample were Krembangan, Rungkut, Bulak, Sambikerep, Gubeng, Sawahan, Tenggilis Mejoyo, Pakal and Wonocolo health centers. Meanwhile, the Out-Sample health center that will be estimated is the Siwalankerto Health Center, Dr. Soetomo, Tanjungsari, Sememi and Tambak Wedi.

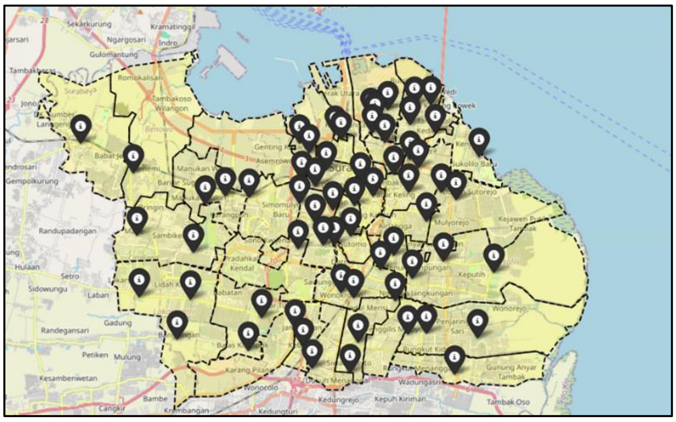

Figure 3 Community Health Center Point Distribution Map for Interpolation

Based on Figure 3, an interpolated map of the distribution of health center locations is obtained. The interpolated health centers include all health centers in Surabaya, including those used for Out-Sample and In-Sample, namely 63 health centers.

C. Assumption Testing

1. *Testing the Spatial Autocorrelation Assumption*

The Spatial Autocorrelation Test used is the Moran's I Test to identify spatial dependencies.

Table 3. Testing the Spatial Autocorrelation Assumption

Variables	Moran's I Index (I)	$Z(I)$	P-Value
ARI Cases in Health Center	0.153	2.301	0.021

Table 3, the Moran's I test statistic is 2.301, where this value is greater than $Z\alpha/2$ namely 1.96. Apart from that, a p-value of 0.021 was obtained, which is smaller than alpha 0.05, so it can be concluded that there is spatial dependence in the data on ARI cases in Surabaya.

2. *Correlation Test*

The primary variable in this research is the number of ARI cases. while PM2.5 concentration is a secondary variable. The correlation test between the two variables was carried out to see the relationship between the two variables. The results of correlation testing between the two research variables can be seen in Table 4.

Table 4. Correlation Test Results Between Variables

Correlation Pearson	Description
0.456	Moderate correlation

Table 4 shows that the two variables are moderately correlated. Thus, these results support the use of the cokriging method in estimating ARI case variables in Surabaya using information from the PM2.5 concentration variable as secondary.

D. Experimental Semivariogram

The process of obtaining the estimated values for $PM_{2.5}$ pollutant levels begins with using the data. The next step is to calculate the distance ($\boldsymbol{h}$) between air monitoring station points in Surabaya and the distance between these station points and the locations to be estimated, using Python software. After determining the distances between station locations, the next step is to divide the distances into k classes. This division is performed using the formula based on Sturges' rule: $\boldsymbol{k = 1 + (3,33 \times \log(n))}$ with $\boldsymbol{n}$ being 81 data points. The output of the distances between station locations and the locations to be estimated, along with the number of classes from the running program. The following step is to calculate the experimental auto covariance and the experimental cross-covariance values using equations (1), (2), and (3). The experimental auto covariance and experimental cross-covariance values can be seen in Table 5 below with the assistance of Python software.

Table 5 Distance Class Division and Experimental Covariance

Class Interval	Distance ($\boldsymbol{h}$)	$\boldsymbol{N}$	Auto Covariance $\boldsymbol{U}$	Auto Covariance $\boldsymbol{V}$	Cross Covariance $\boldsymbol{UV}$
0,000000-0,026538	0,013269	8	0,859097	1.000000	0,404229
0,026538-0,053276	0,039957	16	0,094080	-0,147104	-0,035467
0,053276-0,079915	0,066595	16	0,001511	-0,207012	0,086540
0,079915-0,106553	0,093234	16	0,251689	0,026677	0,226978
0,106553-0,133191	0,119872	12	-0,097254	-0,314024	-0,589158
0,133191-0,159829	0,146510	9	-0,854213	-0,053354	0,320781
0,159829-0,186467	0,173148	2	-2,631829	0,435976	-1,389694
0,186467-0,213105	0,199786	2	-0,300529	-0,396341	-0,360207

With $\boldsymbol{N}$ being the number of pairs at distance h, U as the first variable (number of ARI cases), and V as the second variable ($PM_{2.5}$). After obtaining the experimental auto covariance and experimental cross covariance, the next step is to determine the spherical auto covariance and spherical cross covariance. For this purpose, based on equations (4), (5), (6), estimation of the values of P (nugget effect), Q (Sill), and r (range) is required. The values of P, Q, and r for spherical auto covariance are determined based on the plot of distance against experimental covariance, while the values of P, Q, and r for spherical cross-covariance are determined based on the plot of distance against experimental cross-covariance. The plot of distance against experimental auto covariance for the first variable (number of ARI cases) is shown in Figure 4 below.

DOI: 10.12962/j27213862.v8i1.20512

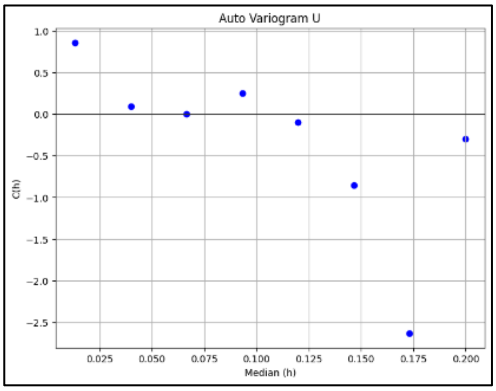

Figure 4 Plot of Distance against Experimental Auto Covariance for the ARI Variable

Based on Figure 4, the plot of distance against experimental auto covariance for the ARI variable, the values obtained are $P = 0.00151$, $Q = 0.859096$, and $r = 0.013269$ for the calculation of spherical auto covariance between ARI variables. The plot of distance against experimental auto covariance for the secondary variable ($PM_{2.5}$) can be seen in Figure 5 below.

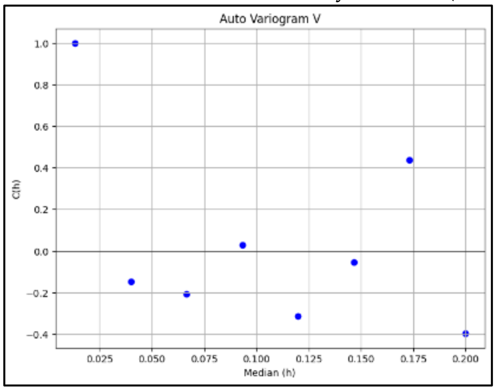

Figure 5 Plot of Distance against Experimental Auto Covariance for the $PM_{2.5}$ Variable

Based on Figure 5, the plot of distance against experimental auto covariance for the $PM_{2.5}$ variable, the values obtained are $P = 0.026676$, $Q = 0.999999$, and $r = 0.013269$ for the calculation of spherical auto covariance between $PM_{2.5}$ variables. The plot of distance against experimental auto covariance for the secondary variable ($PM_{2.5}$) can be seen in Figure 6 below.

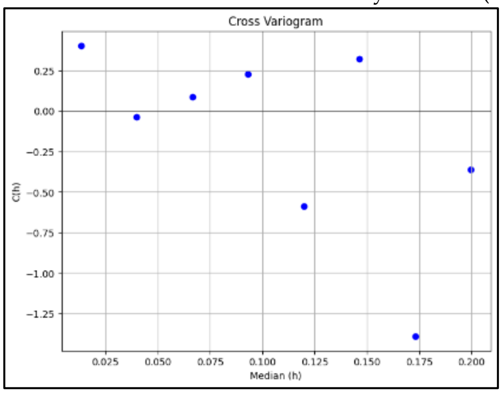

Figure 6 Plot of Distance against Experimental Cross Covariance

Based on Figure 6, the values obtained are $P = 0.086540$, $Q = 0.404229$, and $r = 0.013269$ for the calculation of spherical auto covariance between first variable and second variable.

E. Theoretical Semivariogram

After obtaining the values of P, Q, and r for each covariance, the formula for calculating the spherical model, according to equation (4), is as follows:

The value of spherical auto covariance for the primary ARI variable is:

$$C_U(h) = \begin{cases} (0{,}00151 + 0{,}859096), & 0 = h \\ (0{,}00151 + 0{,}859096)\left(1 - \dfrac{3h}{2(0{,}013)} + \dfrac{h^3}{2(0{,}013)^3}\right), & 0 < h \leq 0{,}013 \\ 0, & h > 0{,}013 \end{cases} \tag{8}$$

The value of spherical auto covariance for the secondary $PM_{2.5}$ variable is:

$$C_V(h) = \begin{cases} (0{,}026676 + 0{,}999999), & 0 = h \\ (0{,}026676 + 0{,}999999)\left(1 - \dfrac{3h}{2(0{,}013)} + \dfrac{h^3}{2(0{,}013)^3}\right), & 0 < h \leq 0{,}013 \\ 0, & h > 0{,}013 \end{cases} \tag{9}$$

The value of spherical cross covariance is:

$$C_V(h)=\begin{cases}(0{,}086540+0{,}404229), & 0=h\\ (0{,}086540+0{,}404229)\left(1-\dfrac{3h}{2(0{,}013)}+\dfrac{h^3}{2(0{,}013)^3}\right), & 0<h\leq 0{,}013\\ 0, & h>0{,}013\end{cases} \tag{10}$$

After obtaining the values of P, Q, and r for each covariance, the formula for calculating the exponential model, according to equation (5), is as follows:

The value of exponential auto covariance for the primary ARI variable is:

$$C(h)=(0{,}00151+0{,}859096)\left[1-exp\left(-\frac{h}{0{,}013}\right)\right] \tag{11}$$

The value of exponential auto covariance for the primary $PM_{2.5}$ variable is:

$$C(h)=(0{,}026676+0{,}999999)\left[1-exp\left(-\frac{h}{0{,}013}\right)\right] \tag{12}$$

The value of exponential cross covariance is:

$$\boldsymbol{C(h)=(0,086540+0,404229)\left[1-exp\left(-\frac{h}{0,013}\right)\right]} \tag{13}$$

After obtaining the values of P, Q, and r for each covariance, the formula for calculating the gaussian model, according to equation (6), is as follows:

The value of gaussian auto covariance for the primary ARI variable is:

$$C(h)=(0{,}00151+0{,}859096)\left[1-exp\left(-\frac{h}{0{,}013}\right)^2\right] \tag{14}$$

The value of gaussian auto covariance for the primary $PM_{2.5}$ variable is:

$$C(h)=(0{,}026676+0{,}999999)\left[1-exp\left(-\frac{h}{0{,}013}\right)^2\right] \tag{15}$$

The value of gaussian cross covariance is:

$$\boldsymbol{C(h)=(0,086540+0,404229)\left[1-exp\left(-\frac{h}{0,013}\right)^2\right]} \tag{16}$$

F. *Comparison of Semivariogram Models on In-Sample Data*

1. Spherical Semivariogram Model

In selecting the semivariogram model on in-sample data using the spherical semivariogram model, the predicted data values are presented in Table 6 below.

Table 6 Prediction Results of In-Sample Data with Spherical Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Krembangan Health Centre	217	214
Rungkut Health Centre	218	215
Bulak Health Centre	424	421
Sambikerep Health Centre	95	92
Gubeng Health Centre	238	235
Sawahan Health Centre	109	106
Tenggilis Mejoyo Health Centre	246	243
Pakal Health Centre	67	64
Wonocolo Health Centre	109	106

By using the spherical semivariogram model, the MSE value on the training data is 14.425 and the MAPE value is 2.761%. Prediction results using this model will produce highly accurate forecasting, because the MAPE value of 2.761% is below 10%.

2. Exponential Semivariogram Model

In selecting the semivariogram model on in-sample data using the exponential semivariogram model, the predicted data values are presented in Table 7 below.

Table 7 Prediction Results of In-Sample Data with Exponential Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Krembangan Health Centre	222	214
Rungkut Health Centre	223	215
Bulak Health Centre	429	421
Sambikerep Health Centre	100	92
Gubeng Health Centre	243	235
Sawahan Health Centre	114	106
Tenggilis Mejoyo Health Centre	251	243
Pakal Health Centre	72	64
Wonocolo Health Centre	114	106

By using the exponential semivariogram model, the MSE value on the training data is 66.408 and the MAPE value is 5.924%. Prediction results using this model will produce highly accurate forecasting, because the MAPE value of 5.924% is below 10%.

3. Gaussian Semivariogram Model

In selecting the semivariogram model on in-sample data using the gaussian semivariogram model, the predicted data values are presented in Table 8 below.

Table 8 Prediction Results of In-Sample Data with Gaussian Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Krembangan Health Centre	229	214
Rungkut Health Centre	230	215
Bulak Health Centre	436	421
Sambikerep Health Centre	107	92
Gubeng Health Centre	250	235
Location	**Prediction Data (z_0)**	**Original Data**
Sawahan Health Centre	121	106
Tenggilis Mejoyo Health Centre	258	243
Pakal Health Centre	79	64
Wonocolo Health Centre	121	106

By using the exponential semivariogram model, the MSE value on the training data is 240.544 and the MAPE value is 11.275%. Prediction results using this model will produce accurate forecasting, because the MAPE value of 11.275% is in the 10-20% interval.

G. Comparison of Semivariogram Models on Out-Sample Data

1. Spherical Semivariogram Model

In selecting the semivariogram model on out-sample data using the spherical semivariogram model, the predicted data values are presented in Table 9 below.

Table 9 Prediction Results of Out-Sample Data with Spherical Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Dr. Soetomo Health Centre	189	165
Sememi Health Centre	189	159
Siwalankerto Health Centre	67	63
Tambak Wedi Health Centre	191	254
Tanjungsari Health Centre	189,	187

By using the spherical semivariogram model, the MSE value on the testing data is 1100.514 and the MAPE value is 13.316%. Prediction results using this model will produce accurate forecasting, because the MAPE value of 13.316% is in the 10-20% interval.

2. Exponential Semivariogram Model

In selecting the semivariogram model on out-sample data using the exponential semivariogram model, the predicted data values are presented in Table 10 below.

Table 10 Prediction Results of Out-Sample Data with Exponential Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Dr. Soetomo Health Centre	177	165
Sememi Health Centre	153	159
Siwalankerto Health Centre	59	63
Tambak Wedi Health Centre	243	254
Tanjungsari Health Centre	186	187

By using the exponential semivariogram model, the MSE value on the training data is 65.896 and the MAPE value is 4.307%. Prediction results using this model will produce highly accurate forecasting, because the MAPE value of 4.307% is below 10%.

3. Gaussian Semivariogram Model

In selecting the semivariogram model on out-sample data using the gaussian semivariogram model, the predicted data values are presented in Table 11 below.

Table 11 Prediction Results of Out-Sample Data with Gaussian Semivariogram Model

Location	Prediction Data (z_0)	Original Data
Dr. Soetomo Health Centre	183	165
Sememi Health Centre	174	159
Siwalankerto Health Centre	62	63
Tambak Wedi Health Centre	223	254
Tanjungsari Health Centre	195	187

By using the exponential semivariogram model, the MSE value on the training data is 312.822 and the MAPE value is 7.587%. Prediction results using this model will produce highly accurate forecasting, because the MAPE value of 7.587% is below 10%.

H. Best Model Selection

To determine the best semivariogram model used for co-kriging interpolation, a comparison of the MAPE values obtained from the prediction of the number of disease cases that occur is used. Table 12 is the result of the MAPE value comparison,

Table 12 The Comparison of the MAPE value on Semivariogram Models

Semivariogram Model	MAPE In-Sample	MAPE Out-Sample	MAPE Overall
Spherical	2.76%	13.32%	8.04%
Esponential	**5.92%**	**4.31%**	**5.11%**
Gaussian	11.27%	7.58%	9.42%

Based on Table 12, of the three types of semivariogram models, the best model is the exponential semivariogram model. Where it has the smallest and most optimal overall MAPE value among other models. The next step is to perform co-kriging interpolation using the exponential semivariogram model.

I. Co-Kriging Interpolation

Based on the best model, which is the exponential semivariogram model, a prediction map of the number of cases of ARI in Surabaya City can be shown in Figure 7 below.

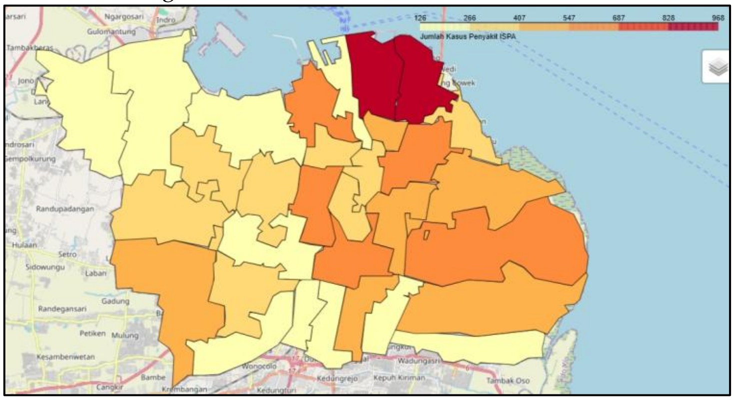

Figure 7 Interpolation Map with Co-Kriging Method

Based on the estimation of the number of ARI cases in 31 sub-districts in Surabaya, the highest number of ARI cases is located in Kenjeran sub-district with a total of 968 ARI patients who seek treatment at the health centre and the lowest number of ARI cases is located in Pakal sub-district with a total of 125 ARI patients who seek treatment at the health centre.

V. CONCLUSIONS AND SUGGESTIONS

This study used 9 community health centres (Puskesmas) in Surabaya as the primary variable ($\boldsymbol{u}$) and 9 $PM_{2.5}$ stations in Surabaya as the secondary variable ($\boldsymbol{v}$) in the co-kriging method to interpolate the number of ARI cases in Surabaya. The selection of the best semivariogram model was made by comparing the MAPE values on Out-Sample data, which included 5 community health centres in Surabaya. Based on the MAPE values, the best semivariogram model was determined to be the Exponential semivariogram model. The MAPE value was 5.11%. This indicates that the Exponential semivariogram model is able to predict the number of ARI cases in Surabaya with high accuracy. Based on the estimation of the number of ARI cases in 31 sub-districts in Surabaya, the highest number of ARI cases is located in Kenjeran sub-district with a total of 968 ARI patients who seek treatment at the health centre and the lowest number of ARI cases is located in Pakal sub-district with a total of 125 ARI patients who seek treatment at the health centre.

Future research is recommended to increase the number of observation points for both the primary variable (community health centers) and the secondary variable ($PM_{2.5}$ stations) to obtain more accurate and representative estimates. Then, future analyses should consider additional factors that influence the spread of ARI, such as weather conditions, population density, and socio-economic factors.

REFERENCES

[1] J. Miles, "WHO global air quality guidelines," *Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.*, 2021.

[2] IQAir, "World Air Quality Report 2023," *IQAir*, 2023.

[3] E. Widiyana, "174.222 Warga Surabaya Terserang ISPA, 6 Ribu di Antaranya Balita," *Detik Jatim*, Surabaya, Sep. 08, 2023. [Online]. Available: https://www.detik.com/jatim/berita/d-6919600/174-222-warga-surabaya-terserang-ispa-6-ribu-di-antaranya-balita

[4] C. Wang, "The impact of car ownership and public transport usage on cancer screening coverage: Empirical evidence using a spatial analysis in England," *J. Transp. Geogr.*, vol. 56, pp. 15–22, 2016, doi: 10.1016/j.jtrangeo.2016.08.012.

[5] R. S. A. Adelia, "Interpolasi Spasial Metode Co-Kriging Menggunakan Semivariogram Isotropik dan Anisotropik," 2017.

[6] U. Batawen, "Estimasi Sumberdaya Terukur Endapan Bijih Nikel Laterit Menggunakan Metode Geostatistik (Ordinary Co-Kriging) (Studi Kasus: Site Tinanggea PT Ifishdeco Tbk. Kecamatan Tinanggea Kabupaten Konawe Selatan, Sulawesi Tenggara)," 2021.

[7] M. Sherman, *Spatial Statistics and Spatio-Temporal Data*. John Wiley and Sons, Inc, 2011.

[8] A. Hooshmand, M. Delghandi, A. Izadi, and K. A. Aali, "Application of kriging and cokriging in spatial estimation of groundwater quality parameters," *African J. Agric. Res.*, vol. 6, no. 14, pp. 3402–3408, 2011.

[9] T. Saifudin, A. Faiza, L. Puspasari, and Z. 'Ilmatun Nurrohmah, "Estimating the Concentration of No2 With the Cokriging Method in the Capital City of Jakarta," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 4, pp. 1985–1996, 2023, doi: 10.30598/barekengvol17iss4pp1985-1996.

[10] A. Djuraidah, S. Rahardiantoro, and A. Desiwari, "Penerapan Metode Cokriging Dengan Variogram Isotropi Dan Anisotropi Dalam Memprediksi Curah Hujan Bulanan Jawa Barat Application Of Cokriging Methods With Isotropy And Anisotropy Variograms In Prediction Of West Java Monthly Rainfall," *J. Meteorol. Dan Geofis.*, 2019.

[11] S. Arba, "Kosentrasi Respirable Debu Particulate Matter (PM 2,5) Dan Gangguan Kesehatan Pada Masyarakat Di Pemukiman Sekitar PLTU," *Promot. J. Kesehat. Masy.*, vol. 9, no. 2, pp. 178–184, 2019.

[12] A. T. Arianto, K. Parmikanti, B. Suhandi, and B. N. Ruchjana, "Peramalan Konsentrasi Particulate Matter 2.5 (PM2.5) menggunakan Model Vector Autoregressive dengan Metode Maximum Likelihood Estimation," *KUBIK J. Publ. Ilm. Mat.*, vol. 5, no. 1, pp. 1–12, 2021.

[13] N. Ho *et al.*, "Retrospective analysis assessing the spatial and temporal distribution of paediatric acute respiratory tract infections in Ho Chi Minh City, Vietnam," 2018.

[14] M. Thomas and P. Bomar, "Upper Respiratory Tract Infection," 2024.

[15] Rahmawati, *Gangguan pernafasan pada anak: ISPA*. Yogyakarta: Nurha Medika, 2012.

[16] L. Lutpiatina, L. Sulistyorini, and H. Notobroto, "Acute Respiratory Infections Associated with Exposure to Biomass Cooking Fuels and Cigarette Smoke among Children Under Five Years of Age in Developing Countries," 2022.

[17] E. L. W. Choo, A. Janhavi, J. R. Koo, S. H. L. Yim, B. L. Dickens, and J. T. Lim, "Association between ambient air pollutants and upper respiratory tract infection and pneumonia disease burden in Thailand from 2000 to 2022: a high frequency ecological analysis," *BMC Infect. Dis.*, vol. 23, no. 1, 2023, doi: 10.1186/s12879-023-08185-0.

[18] N. A. C. Cressie, *Statistics for Spatial Data. Computational Statistics & Data Analysis*. John Wiley and Sons, Inc, 1993.

[19] A. Puntodewo, S. Dewi, and J. Tarigan, *Sistem Informasi Geografis Untuk Pengolahan Sumber Daya Alam*. 2023.

[20] D. Nerini, P. Monestiez, and C. Manté, "Cokriging for spatial functional data," *J. Multivar. Anal.*, vol. 101, no. 2, pp. 409–418, 2010, doi: 10.1016/j.jmva.2009.03.005.

[21] P. J. Curran, "The semivariogram in remote sensing: An introduction," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 493–507, 1988, doi: 10.1016/0034-4257(88)90021-1.

[22] A. B. McBratney and R. Webster, "Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates," *J. Soil Sci.*, vol. 37, no. 4, pp. 617–639, 1986, doi: 10.1111/j.1365-2389.1986.tb00392.x.

[23] H. B. Isaaks and R. M. Srivastava, *An Introduction to Applied Geostatistics*. 1989. Accessed: May 31, 2024. [Online]. Available: https://books.google.com/books/about/Applied_Geostatistics.html?id=vC2dcXFLI3YC