International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 7 Issue 3, March - 2023, Pages: 13-18

Prediction of Carbon Monoxide Pollutant Levels Average in DKI Jakarta Province Using the Ordinary Kriging Method

Syehvira Shal Shabilla[1], Toha Saifudin[2*], Nur Chamidah[3], M. Fariz Fadillah Mardianto[4], Siti Maghfirotul Ulyah[5]

[1,2,3,4,5]Statistics Study Program, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[1]syehvira.al.abilla-2019@fst.unair.ac.id
[2]Corresponding author: tohasaifudin@fst.unair.ac.id
[3]nur-c@fst.unair.ac.id
[4]m.fariz.fadillah.m@fst.unair.ac.id
[5]maghfirotul.ulyah@fst.unair.ac.id

Abstract: *DKI Jakarta is the most densely polluted city in Indonesia according to the 2021 IQ Air report. The main source of pollution in DKI Jakarta is caused by transportation emissions. Even though transportation emissions are the largest sector producing CO pollutants, it even reaches 70.5%. The pollutant carbon monoxide (CO) is also the most dangerous pollutant according to research by the New Jersey Department of Environmental Protection. On the other, DKI Jakarta only has five monitoring points for CO levels. Therefore, predictions are made at location points that are not sampled to better represent the condition of CO pollutant levels in DKI Jakarta. This study aims to provide information to the public and the government to immediately take preventive and anticipatory steps. The results showed that the best prediction of CO pollutant levels using the Ordinary Kriging method was exponential with a Mean Absolute Percentage Error (MAPE) of 10.001% which means it has good accuracy. The prediction results for the 44 districts as a whole are within the threshold. There are 26 sub-districts included in the unhealthy for sensitive groups category and 18 sub-districts included in the category. Thus, the government still has to pay more attention to optimizing the reduction of air pollution, especially due to CO pollutants.*

Keywords— carbon monoxide; DKI Jakarta; ordinary kriging; prediction; thematic map

1. INTRODUCTION

Air pollution is a very crucial global problem. The World Health Organization (WHO) states that more than 6000 cities in 117 countries show high levels of air pollution and health risks. About 99% of the global population breathes air that exceeds WHO limits [1]. This indirectly threatens the achievement of the Sustainable Development Goals (SDGs), precisely the third goal with the target of reducing the number of deaths and morbidity, one of which is due to air pollution [2].

DKI Jakarta is the center of government and various activities. This resulted in the high mobility of people in the transportation sector. According to [3], the transportation sector contributes 75% as a source of air pollution. One of the most significant pollutants produced from transportation emissions is carbon monoxide with a contribution of 70.5% [4]. CO pollutant is a gas that is dangerous if inhaled beyond safe limits. This is because CO is able to bind to hemoglobin 210 times greater than oxygen [5]. Meanwhile, according to the New Jersey Department of Environmental Protection, the safe threshold for CO pollutant levels is 20 parts per million (ppm).

CO pollutant levels in DKI Jakarta are routinely monitored at five monitoring stations, including Bundaran HI, Jagakarsa, Kelapa Gading, Kebun Jeruk, and Lubang Buaya [6]. The level of CO pollutants in DKI Jakarta at the five monitoring station points is still within the threshold but varies greatly between locations. The CO pollutant level average in 2021 at the five monitoring points is 12.003. Even though it is below the threshold, according to the New Jersey Department of Environmental Protection, the range of 9.5-12.4 ppm is included in the unhealthy category for sensitive groups [7]. On the other hand, administratively DKI Jakarta Province is a large area consisting of one district, five cities, and 44 sub-districts. In order to be able to represent CO pollutant levels in DKI Jakarta Province with limited sample points, a method is needed that can predict CO pollutant levels at location points that are not sampled. One of the analyzes that can be used is the Kriging analysis.

Kriging is a method of interpolating between data points based on sampled points within a certain range which produces unbiased linear estimates [8]. There are several types of Kriging methods, among others Ordinary Kriging, Universal Kriging, and Cokriging [9]. Ordinary Kriging is the simplest kriging method in geostatistics and is used when the population means are not known. Ordinary Kriging is able to estimate variable values at locations that are not sampled by weighting similar data at other locations and are able to produce an estimator that is the Best Linear Unbiased Estimator [10].

Several similar studies related to Ordinary Kriging include research by [11] regarding the prediction of PM 2.5 concentrations in Surabaya using Ordinary Kriging showing high accuracy with a Mean Absolute Percentage Error (MAPE) value of 5.6%. Research by [12] regarding the estimation of NO_2 concentrations in the city of Yogyakarta, showed that Ordinary Kriging was more accurate than the Inverse Distance Weighted method with Root Mean Square Error (RMSE) values of 0.4847 and 0.5224, respectively. Research by [13] regarding the prediction of Cu concentrations in the Chehlkureh deposit, SE Iran states that Ordinary Kriging is more accurate than Simple Kriging. Research by [14] regarding the estimation of CO gas concentrations at several locations in Semarang City with Ordinary Kriging showed high accuracy with MAPE of 4.6%.

Based on these facts, it is necessary to conduct research to predict CO pollutant levels in each unsampled sub-district, namely as many as 44 sub-districts using the Ordinary Kriging method. This study aims to provide information on CO pollutant levels at unsampled location points and as a consideration for the Regional Government of DKI Jakarta in making policies, especially as a preventive and anticipatory measure in dealing with air pollution in DKI Jakarta in particular.

2. RESEARCH METHODS

2.1 DATA SOURCE

This study uses secondary data in the form of average levels of carbon monoxide pollutant in DKI Jakarta Province in 2021. This data is obtained from a collection of data related to the Air Pollution Standard Index (ISPU) on the official website of the DKI Jakarta Provincial Government's Integrated Data Portal, namely data.jakarta.go.id.

2.2 VARIABLE

The research variable used is the average of CO pollutant level in 2021 in DKI Jakarta. There are three parameters that play a role, namely (u, v) as the latitude and longitude coordinates of the measurement locations and Z as the average value of CO pollutant levels in 2021 with ppm units in five locations including Bundaran HI, Jagakarsa, Kelapa Gading, Kebun Jeruk, and Lubang Buaya.

2.3 STEP OF THE RESEARCH

The steps to achieve the objectives of this study are as follows:

1. Conduct a descriptive statistical analysis of data on CO pollutant levels for 2021 at the five monitoring points in DKI Jakarta.
2. Perform performance analysis of the Ordinary Kriging method with Spherical, Exponential, and Gaussian models for the five monitoring points alternately.
   a. Specify a prediction location $u_0 = (x_0, y_0)$. Determination of one location of prediction is carried out alternately at five monitoring points, provided that the location of the prediction is not the four observation locations $u_i = (x_i, y_i)$ with $i = 1, 2, \dots, 5$.
   b. Enter the latitude and longitude coordinate data for the observation location $u_i = (x_i, y_i)$, prediction location $u_0 = (x_0, y_0)$, and the CO pollutant level value as $Z(u_i)$.
   c. Calculates the Euclidian distance between observation locations. The distance between the i-th location with coordinates (x_i, y_i) to the j-th location with (x_j, y_j) so that the Euclidian distance can be calculated with the following formula:
   $$h_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \quad (1)$$
   with $i, j = 1, \dots, n$
   d. Calculate the Euclidean distance between the observation location and the prediction location using the following formula:
   $$h_{i0} = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \quad (2)$$
   with $i = 1, \dots, n$
   e. Calculating the experimental semivariogram with the following formula [15]:
   $$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(u_i) - Z(u_i + h)]^2 \quad (3)$$
   with $\gamma(h)$ is a semivariogram with a distance h, $Z(u_i)$ and $Z(u_i + h)$ are the observed values for CO pollutant levels at location i-th and location i-th with the addition of h distance, and $N(h)$ is the number of data pairs that have h distance.
   f. Determine the range (a) and sill $(C_0 + C)$. The range is the distance when the semivariogram reaches a stable period or sill, while the sill value is equal to the variance of the data.
   g. Calculate the theoretical semivariogram of the Spherical, Exponential, and Gaussian models using the following formula [16]:
   Spherical Models:
   $$\gamma(h) = \begin{cases} [C_0 + C]\left[\left(\frac{3h}{2a}\right) - 0.5\left(\frac{h}{a}\right)^3\right], & h \le a \\ [C_0 + C] & , h > a \end{cases} \quad (4)$$
   Exponential Models:
   $$\gamma(h) = [C_0 + C]\left[1 - \exp\left(-\frac{h}{a}\right)\right] \quad (5)$$
   Gaussian Models:
   $$\gamma(h) = [C_0 + C]\left[1 - \exp\left(-\frac{h^2}{a^2}\right)\right] \quad (6)$$
   h. Make plots of distances between locations with experimental semivariograms and theoretical semivariograms.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 7 Issue 3, March - 2023, Pages: 13-18

i. Forming a matrix $\boldsymbol{P}$ which contains the theoretical semivariograms between observation locations and $\boldsymbol{S}$ which contains the theoretical semivariograms of the observation locations and the prediction locations. Then calculate the weighting value of each observation location with equation [16]:

$$\boldsymbol{Q} = \boldsymbol{P}^{-1}\boldsymbol{S} \quad (7)$$

The following shows in detail the matrix elements $\boldsymbol{P}$, $\boldsymbol{S}$, and $\boldsymbol{Q}$:

$$\boldsymbol{P} = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 \\ \vdots & \vdots & & \vdots & 1 \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix},$$

$$\boldsymbol{S} = \begin{bmatrix} \gamma(h_{10}) \\ \gamma(h_{20}) \\ \vdots \\ \gamma(h_{n0}) \\ 1 \end{bmatrix}, \text{ dan } \boldsymbol{Q} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix}$$

j. Calculate the predicted value with the following formula [16]:

$$\hat{Z}(u_0) = \sum_{i=1}^{n} w_i Z(u_i) \text{ where } \sum_{i=1}^{n} w_i = 1 \quad (8)$$

k. Calculating the MAPE value for each model with the following formula [17]:

$$MAPE = \frac{\sum_{k=1}^{n}\left|\frac{Y_k - \hat{Y}_k}{Y_k}\right|}{n} \times 100\% \text{ , } k = 1, \ldots, 5 \quad (9)$$

where n is the amount of data, Y_k as the actual data on the k-th and $\hat{Y}_k$ as the predicted data on the k-th observation.

l. Interpret MAPE based on the following categories [18]:

Table 1. MAPE Category

MAPE	Interpretation
< 10%	*Highly Accuracy*
10% - 20%	*Good Accuracy*
20% - 50%	*Reasonable Accuracy*
>50%	*Bad Accuracy*

3. Predict CO pollutant levels in each of the 44 sub-districts in DKI Jakarta using the best theoretical semivariogram model based on the smallest MAPE value using the same steps as in step 2.
4. Make a thematic map using ArcMap Software.

3. RESULT AND DISCUCCION

3.1 Descriptive Statistics

The average of CO pollutant levels at the five monitoring points, namely Bundaran HI, Jagakarsa, Kelapa Gading, Kebun Jeruk, and Lubang Buaya in 2021 are spread as shown in the plot below:

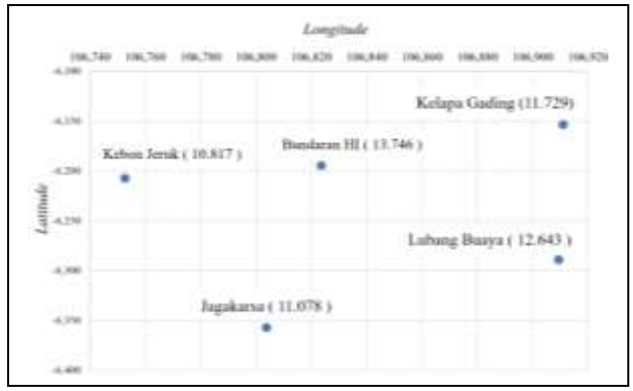

Fig. 1. Distribution Plot of CO Pollutant Level Monitoring Point Locations

The longitude (X) axis describes the east longitude area, while the latitude (Y) axis describes the south latitude. Based Figure 1, the monitoring point for Bundaran HI representing the city center is in the middle position, Kelapa Gading is in the northernmost position representing the North Jakarta area, Lubang Buaya is in the easternmost position representing North Jakarta, Jagakarsa is in the southernmost position representing South Jakarta, dan Kebun Jeruk is in the westernmost position representing West Jakarta. The descriptive statistical analysis of CO pollutant levels at the five monitoring points is as follows:

Table 2. Descriptive Statistics

Mean	**Standard deviation**	**Variance**	**Maximum**	**Minimum**
12.003	1.203	1.446	13.746	10.817

Based on Table 2, the average of CO pollutant level at the five monitoring points in 2021 was 12.003 ppm, the standard deviation was 1.203, and the variance was 1.446. The maximum average of CO pollutant level is 13.746 at the Bundaran HI and a minimum of 10.817 at the Kebon Jeruk monitoring point.

3.2 Prediction of CO Pollutant Levels at the Five Monitoring Points

The location of the estimation used is five monitoring points which are carried out alternately, so that the observation locations used are only four other monitoring points. Based on the results of the experimental semivariogram calculation of CO pollutant levels at the Bundaran HI monitoring point, the distance between the observation points and their semivariogram values is obtained as in the following table.

Table 3. Experimental Semivariogram

No	Distance (h_{ij})	Observation Value at the point S_i $(Z(S_i))$	Observation Value at the point $S_i + h$ $(Z(S_i + h))$	Experimental Semivariogram $(\gamma(h))$
1	0.126	11.078	12.643	1.225
2	0.135	11.729	12.643	0.417
3	0.158	11.078	10.817	0.034
4	0.167	11.729	10.817	0.416
5	0.177	12.643	10.817	1.668
6	0.230	11.729	11.078	0.212

Based on Table 3, six pairs of distances between observation points were obtained as well as experimental semivariogram values at each distance. After calculating the experimental semivariogram, the sill parameter value is 0.341 and the range is 0.147. Sill values and ranges are used to calculate theoretical semivariograms for Spherical, Exponential, and Gaussian models. The results of the theoretical semivariogram calculations for the three models are presented in the following figure.

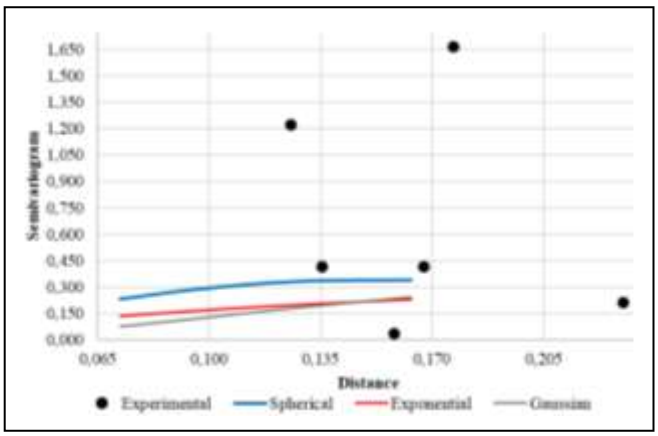

Fig. 2. Experimental Semivariogram Plots with Theoretical Semivariogram Spherical, Exponential, and Gaussian Models

So that the Spherical, Exponential, and Gaussian model formulas are obtained as follows for the Bundaran HI:

1. Spherical Models

$$\gamma(h) = \begin{cases} 0.341\left[\left(\frac{3h}{2(0.147)}\right) - 0.5\left(\frac{h}{0.147}\right)^3\right], & h \leq 0.147 \\ 0.341 & , \ h > 0.147 \end{cases}$$

2. Exponential Models

$$\gamma(h) = 0.341\left[1 - \exp\left(-\frac{h}{0.147}\right)\right]$$

3. Gaussian Models

$$\gamma(h) = 0.341\left[1 - \exp\left(-\frac{h^2}{(0.147)^2}\right)\right]$$

The same steps were carried out for the other four estimating points. After obtaining the three model formulas, then predictions are made on CO pollutant levels at the five prediction points with the three models. The predicted data has known sample locations and it is assumed that these locations are not sampled. Then the best model is selected based on the smallest MAPE. The following is the prediction of CO pollutant levels at five monitoring points using the three models:

Table 4. Prediction Results of the Five Monitoring Points with Spherical, Exponential, and Gaussian Models

Location	Actual Data	Spherical	Exponential	Gaussian
Bundaran HI	13.746	11.372	12.686	11.356
Kelapa Gading	11.729	12.062	12.686	13.192
Jagakarsa	11.078	12.220	12.103	11.589
Lubang Buaya	12.643	11.836	11.819	11.758
Kebon Jeruk	10.817	12.533	12.803	13.362
MAPE (%)		10.535	10.001	13.001

Based on Table 4, it is obtained that the smallest MAPE is 10.001 in the Exponential model, so that the Exponential model is chosen to be the best model for predicting CO pollutant levels in each subdistrict.

3.3 Prediction of CO Pollutant Levels in Each District

Prediction of CO pollutant levels in each sub-district was carried out using the Ordinary Kriging Exponential model method. With the same work steps, the prediction results for 44 sub-districts in DKI Jakarta Province are obtained as follows:

Table 5. Prediction Results of CO Pollutant Levels in 44 Districts in DKI Jakarta Province

Subdistrict	Prediction Results	Subdistrict	Prediction Results
Cempaka Putih	12.676	Cilandak	11.636
Gambir	13.023	Jagakarsa	11.486
Johar Baru	12.958	Kebayoran Baru	11.954
Kemayoran	12.514	Kebayoran Lama	11.698
Menteng	13.333	Mampang Prapatan	12.433
Sawah Besar	12.693	Pancoran	12.654
Senen	13.225	Pasar Minggu	12.133
Tanah Abang	13.301	Pesanggrahan	11.369
Cilincing	11.749	Setiabudi	13.313
Kelapa Gading	11.896	Tebet	12.861
Koja	11.813	Cakung	11.985
Pademangan	12.418	Cipayung	12.123

Penjaringan	12.156	Ciracas	12.172
Tanjung Priok	11.924	Duren Sawit	12.424
Cengkareng	11.462	Jatinegara	12.734
Grogol Petamburan	12.171	Kramat jati	12.408
Kali Deres	11.437	Makasar	12.487
Kebun Jeruk	11.358	Matraman	12.821
Kembangan	11.019	Pasar Rebo	11.898
Palmerah	12.211	Pulo Gadung	12.346
Taman Sari	12.391	Kepulauan Seribu Utara	11.698
Tambora	12.235	Kepulauan Seribu Selatan	11.694

Based on the prediction results in Table 5, it can be visualized in a thematic map as follows:

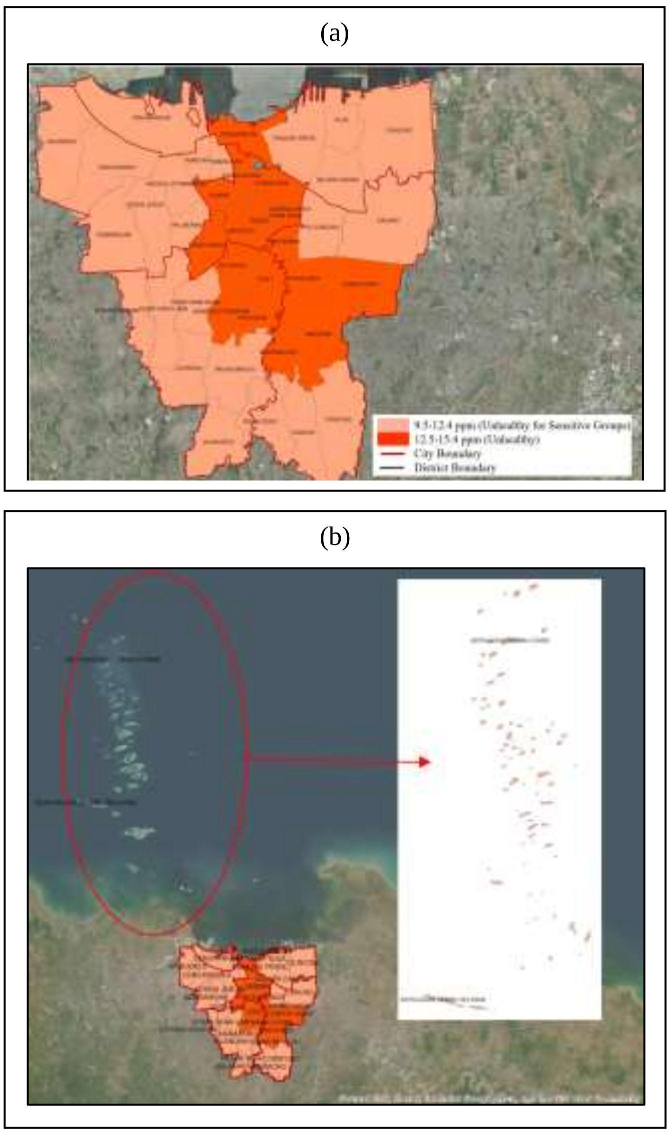

Fig. 3. Thematic Map (a) Central, West, South and East Jakarta Areas, (b) Kepulauan Seribu Regency

Based on the thematic map in Figure 3, there are 26 sub-districts showing predicted results of CO pollutant levels of more than 12.4 ppm and 18 sub-districts below 12.4 ppm. Based on the categorization by the New Jersey Department of Environmental Protection of 9.5-12.4 ppm, it is included in the unhealthy category for sensitive groups, meaning that it should limit prolonged outdoor activity to groups of children, active adults, and people with respiratory diseases such as asthma. The other 18 sub-districts are included in the unhealthy category, namely in the range of 12.5-15.4 ppm, meaning that everyone should limit prolonged outdoor activities [7].

4. Conclusion

Research shows that the Ordinary Kriging method with the Exponential model has the best performance in predicting CO pollutant levels in DKI Jakarta with a MAPE of 10.001% or included in the high accuracy category. Meanwhile overall, the average of CO pollutant level is below the safe threshold. There were 26 sub-districts showing results of predictions of CO pollutant levels classified as unhealthy for sensitive groups and 18 other sub-districts included in the unhealthy category. Therefore, the Regional Government of DKI Jakarta is still expected to take preventive measures to deal with air pollution, especially to reduce CO pollutant levels.

5. References

[1] WHO. (2022). Billions of People Still Breathe Unhealthy Air: New WHO Data. https://www.who.int/news/item/04-04-2022-billions-of-people-still-breathe-unhealthy-air-new-who-data, Retrieved 6 November 2022.

[2] Kementerian PPN. (2017). Terjemahan Tujuan & Target Global Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs). Jakarta: Kementerian PPN/Bappenas.

[3] Kautsar, M. F., & Herlinda. (2021). Air Pollution CISDI Report 2021. Probolinggo: Center for Indonesia's Strategic Development Initiatives (CISDI).

[4] Sasmita, A., Reza, M., Elystia, S., & Adriana, S. (2022). Analisis Pengaruh Kecepatan dan Volume Kendaraan Terhadap Emisi dan Konsentrasi Karbon Monoksida di Jalan Jenderal Sudirman, Kota Pekanbaru. Jurnal Teknik Sipil, 16(4), pp. 269-279.

[5] Rifai, M. H., Rachmat, H., & Prasetyo, M. D. (2022). Pemanfaatan Internet of Things (IOT) untuk Rancang Bangun UAV (Unmanned Aerial Vehicle) Alat Pengukur Polutan CO dan CO2 di Pabrik Manufaktur Menggunakan ESP-NOW. eProceedings of Engineering, 8(5), pp. 7096-7106.

[6] Nelvidawati, N., & Roza, A. (2022). Perbandingan Nilai ISPU Harian di 5 Titik Pemantauan Udara DKI Jakarta pada Saat Lockdown Maret 2020. Construction And Material Journal, 4(1), pp. 23-30.

[7] New Jersey Department of Environmental Protection. (2021). 2021 New Jersey Air Quality Report. Trenton, New Jersey: New Jersey Department of Environmental Protection.

[8] Maulana, I. A., Prasetiyowati, S. S., & Sibaroni, Y. (2022). Prediction Of Rainfall Classification Of Java Island with Ann-Feature Expansion and Ordinary Kriging. Jurnal Media Informatika Budidarma, 6(4), pp. 1988-1997.

[9] Baffoe-Twum, E., Asa, E., & Awuku, B. (2022). Estimation of Annual Average Daily Traffic (AADT) Data for Low-Volume Roads: A Systematic Literature Review and Meta-Analysis. Emerald Open Research, 4(13), pp. 1- 22.

[10] Khaq, M. N., Bargawa, W. S., & Winarno, E. (2022). Estimasi Sumberdaya Endapan Nikel Laterit Sulawesi Tenggara dengan Metode Ordinary Kriging. Jurnal Sumberdaya Bumi Berkelanjutan (SEMITAN), V1(1), pp. 1-6.

[11] Fitri, D. W., Afifah, N., Anggarani, S. M. D., & Chamidah, N. (2021). Prediction Concentration of PM2.5 in Surabaya Using Ordinary Kriging Method. In AIP Conference Proceedings (Vol. 2329). American Institute of Physics Inc. https://doi.org/10.1063/5.0042284

[12] Safira, M. C., Fauzan, A., & Adhiwibawa, M. A. S. (2022). Interpolasi Polutan Nitrogen Dioksida (No2) Dengan Pendekatan Ordinary Kriging Dan Inverse Distance Weighted (Studi Kasus Di Kota Yogyakarta). Jurnal Aplikasi Statistika & Komputasi Statistik, 14(2), pp. 55-66.

[13] Daya, A. A., & Bejari, H. (2015). A Comparative Study Between Simple Kriging and Ordinary Kriging for Estimating and Modeling The Cu Concentration in Chehlkureh Deposit, Se Iran. Arabian Journal of Geosciences, vol. 8, pp. 6003-6020.

[14] Nugroho, F. W., Suryono, S., & Suseno, J. E. (2019). Fog Computing for Monitoring of Various Area Mapping Pollution Carbon Monoxide (CO) with Ordinary Kriging Method. In 2019 Fourth International Conference on Informatics and Computing (ICIC), vol. 2019, pp. 1-6.

[15] Cressie, N. A. C. (1993). Statistics for Spatial Data. New York: John Wiley and Sons, Inc.

[16] Issaks, E. H., & Srivastava, R. M. (1989). Applied Geostatistics. New York: Oxford University Press, Inc.

[17] Hendriani, T., Yamin, M., & Dewi, A. P. (2016). Sistem Peramalan Persediaan Obat dengan Metode Weight Moving Average dan Reorder Point (Studi Kasus: Puskesmas Soropia). Semantik, 2(2), pp. 207-214.

[18] Lewis, C. D. (1982). Industrial and Business Forecasting Methods. London: Butterworth-Heinemann.