ANALYSIS OF KRIGING INTERPOLATION MODELS FOR PARTICULATE MATTER LEVELS IN THE NATIONAL CAPITAL REGION, PHILIPPINES

Rodyvito Angelo Torres [1 *], Roseanne Ramos [1, 2], Bernadette Anne Recto [1], Ayin Tamondong [1, 2]
[1]Training Center for Applied Geodesy and Photogrammetry, University of the Philippines
Diliman, Quezon City, Philippines
Email: rbtorres@alum.up.edu.ph, bbrecto@up.edu.ph

[2] Department of Geodetic Engineering, University of the Philippines
Diliman, Quezon City, Philippines
Email: rvramos@up.edu.ph, amtamondong@up.edu.ph

KEY WORDS: Air Quality, Kriging Interpolation, Particulate Matter

ABSTRACT: The estimation of air quality concentrations using remote sensing and Geographic Information Systems relies on the spatial resolution of the satellite images for the smallest area it can measure. Spatial interpolation provides a way to generate maps of higher spatial resolution with accuracy based on spatial autocorrelation. The study aims to test Kriging interpolation techniques and models applied on different image resolutions to determine which interpolation method resulted in the highest accuracy when compared to 8 to 13 ground monitoring station data for $PM_{2.5}$ and PM_{10}, respectively, for 2019. Kriging interpolation techniques, specifically Ordinary Kriging, Simple Kriging, and Universal Kriging were tested with different models: Stable, Circular, Spherical, Exponential, and Gaussian. Fifteen (15) models each for fine particulate matter ($PM_{2.5}$) and coarse particulate matter (PM_{10}) were evaluated using statistical analysis. $PM_{2.5}$ interpolation models showed R^2 values from 0.76 to 0.92 while PM_{10} models showed R^2 values from 0.19 to 0.39. Ordinary Kriging with Spherical model showed the best results for interpolated maps with spatial resolutions from 5 to 100 meters for $PM_{2.5}$ ($R^2 = 0.80 – 0.83$) while Simple Kriging - Circular model for PM_{10} ($R^2 = 0.24$). For spatial resolutions of 100 to 1000 meters, Simple and Ordinary Kriging with Gaussian Model showed the best results ($R^2 = 0.28 – 0.39$) for PM_{10}. PM maps were generated from the best models and showed areas of high particulate matter concentrations within Manila and Navotas fish ports in the National Capital Region. Researchers can further explore other interpolation models with longer observation periods to improve the study.

1. INTRODUCTION

Air pollution is still one of the main concern in the Philippines, causing environmental harm and health hazards to the public, especially the vulnerable ones. Specifically, a research showed that 66,230 deaths was primarily caused by air pollution and equated to a total economic cost of 11.9% of the GDP of the Philippines in 2019 (Myllyvirta et. al., 2023). With the global need to measure air pollution and manage air quality, scientists are utilizing technologies to monitor air pollution frequently in large scales. Specifically, researchers are using and modifying satellite system and images, such as GOME-1 ERS-2, MODIS, ENVISAT, Aura OMI, and Sentinel-5P TROPOMI, to measure air quality parameters and pollutants (Abdel-Sattar, 2019). The use of satellite images enables researchers to analyse air quality in high temporal resolutions, however, the images' spatial resolution limit how fine the analysis researchers can perform over a study area. Spatial interpolation is a technique to estimate unknown values at specific points based on the measurement of points with known values. Nearest neighbor is the simplest interpolation technique where unknown values are determined by the value of the nearest pixel. According to a study, the most commonly used interpolation methods are inverse distance weighted (IDW) and kriging (Gong et. al., 2014). Inverse Distance Weighting (IDW) interpolation calculates an average value based on the distance of the known points to the unknown point. Specifically, IDW computes the value of the unknown point by getting the sum of the ratio of the observed values over their distance to the unknown point raised to a power set by the user depending on how much emphasis on the known values the user wants. On the other hand, Kriging interpolation is a type of resampling technique used to sharpen images and fill missing data through estimation based on spatial data. Kriging Interpolation produces prediction surfaces together with measures of accuracy through geostatistical methods. Specifically, it interpolates by graphing out the variance of all pairs of data in terms of distance through a semi-variogram following the concept of Tobler's First Law of Geography where closer things are more related than distant things. In 1991 a researcher used and tested kriging analysis for digital satellite imagery as opposed to usual interpolation methods such as nearest-neighbor, bilinear interpolation, and cubic convolution where either the technique produces block-like images or pixel values vastly different from the original value, and determined that kriging interpolation performed better than cubic convolution when enlarging image (McGee, 1991). Another study used the aforementioned

methods for analyzing microscopic chemical changes and found that Ordinary Kriging method retained the original features better than IDW (Milillo & Gardella, 2012). For the purpose of radioactive contamination or estimating pollution, a pair of researchers conducted a study wherein Ordinary Kriging results were better than that of IDW's (Mabit and Bernard, 2017). Another study mentioned that the geostatistical kriging interpolation technique proves to be a viable method for predicting the Annual Average Daily Traffic (AADT) values for roads with no available data (Shamo et al., 2012). Moreover, the paper explored various kriging types such as simple kriging, ordinary kriging, and universal kriging with different variogram models such as nugget effect, spherical, exponential, Gaussian, and power on AADT data of Washington State. The results have shown that ordinary or universal kriging combined with the exponential variogram option were the most appropriate interpolation method. A more recent study used ordinary kriging and inverse distance weighting to assess and map nitrogen dioxide (NO_2), sulphur dioxide (SO_2), and coarse particulate matter (PM_{10}) in Rajasthan, India (Kamboj et al., 2022).Specifically, the results showed significant agreement between predicted values from inverse distance weighting and ordinary kriging and monitored air quality data. Researchers were able to determine PM_{10} as the leading contributor to air quality index each season with NO_2 and SO_2 concentration levels below the standard allowable level With these findings from these previous studies, the researchers aim to test Kriging interpolation techniques and models applied on different image resolutions to determine which interpolation method resulted in the highest accuracy. The modeling tests are applied relative to the ground monitoring station datasets obtained in 2019 for $PM_{2.5}$ (in 8 stations) and PM_{10} (in 13 stations) in the National Capital Region, Philippines.

2. METHODOLOGY

2.1 Study Area

The National Capital Region (NCR) is a highly populated and urbanized area and serves as the center of government and economy of the Philippines. Figure 1 shows the National Capital Region boundary and PM monitoring stations over a base map from © Microsoft Bing VirtualEarth (2023). The region is divided into 16 cities and 1 municipality categorized into four districts with a total land area of around 620 square kilometres. The region is classified under Climate Type 1 according to Coronas classification. Climate Type 1 areas have two distinct seasons: (1) dry season from November to April, and (2) wet season during the rest of the year, with months of June to September as maximum rain period (PAGASA, 2014; Tolentino et. al., 2016). Mobile sources were identified from previous research projects as the top contributor for $PM_{2.5}$ in the region and were observed to be increasing when closer to major roads. Specifically, the primary and secondary aerosols coming from local vehicular emissions affect the region's $PM_{2.5}$ mass concentration (Bagtasa & Yuan, 2020).

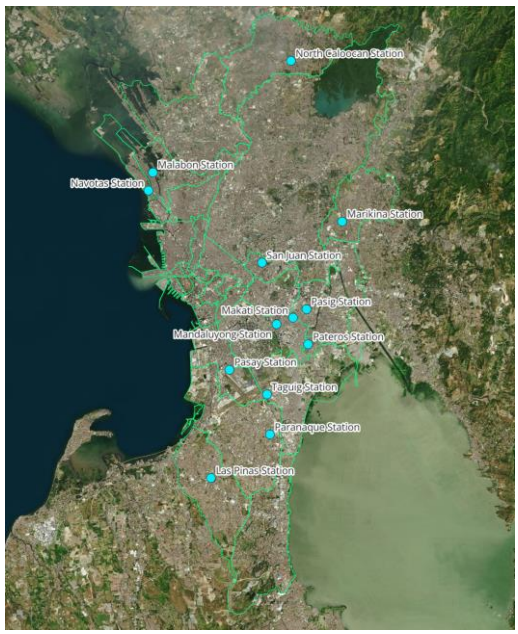

Figure 1. The National Capital Region and PM monitoring stations of the Philippines (base map: Microsoft Bing VirtualEarth, 2023).

2.2 Air Quality Monitoring Stations

The Department of Environment and Natural Resources – Environmental Management Bureau (DENR-EMB) is the regulatory agency dictated by the government to measure and monitor air quality in the Philippines. The agency measures air quality pollutant concentrations such as fine and coarse particulate matter, NO_2, and SO_2 through manual surveys on

selected sampling sites and continuous monitoring stations based on the National Ambient Air Quality Standards (NAAQS). For the study, eight and thirteen observation points for the ground fine $PM_{2.5}$ and coarse PM_{10} data, respectively, were gathered from the Continuous Air Monitoring Stations (CAMS). Yearly aggregate for each station was computed from the hourly ground monitoring station data for 2019. The yearly aggregate was then used to validate the resulting concentration from the interpolated maps.

2.3 Kriging Interpolation

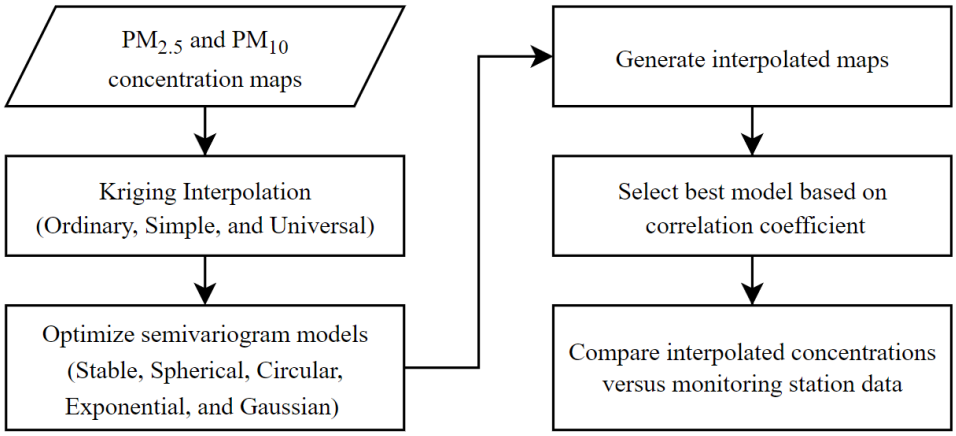

Figure 2. General methodology for the interpolation of PM maps using Kriging Interpolation.

Figure 2 shows the general flowchart of Kriging Interpolation process for PM maps. The PM point vector files generated from a LUR model were used as input files for the Kriging Interpolation. Kriging is an interpolation method in which weights are calculated in a way that nearby points are given more weight than those farther away while also generating uncertainty values of the interpolated pixel surroundings. Moreover, clustering of points is also considered where cluster of points is weighted less heavily to reduce bias in prediction. In this study, three types of Kriging Interpolation were tested, Ordinary Kriging, Simple Kriging, and Universal Kriging. Ordinary Kriging is one of the simplest type of kriging with the assumption of a constant unknown, while Simple Kriging assumes that the constant mean is known and is stricter on the assumption of second-order stationarity. Moreover, in Simple Kriging, the known mean is first subtracted from the data and then later added back after the estimation of residuals. This makes Ordinary Kriging more robust even with some discrepancy from the original assumption. Lastly, Universal Kriging allows differing mean of the values in deterministic ways for different locations still defined by a calculated spatial trend, which can be a linear or polynomial function. In general, the calculation of weights applied to the sample points determined from the variogram is based on the formula:

$$\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \qquad (1)$$

where $\hat{Z}$ = predicted value
x_0 = location of the predicted point
λ_i = weight for the measured value at sampled point
Z = measured value
x_i = location of measured value
N = number of measured values

After selecting the Kriging type, five semivariogram models were tested: (1) Stable, (2) Circular, (3) Spherical, (4) Exponential, and (5) Gaussian. The semivariogram is used to determine the spatial covariance structure of the sampled points where the weights are derived to interpolate the unsampled points. The semivariogram is a plot of a value based on the mean-squared difference of each pair of sampled data versus the distance between them. The option on what semivariogram model to use is generally user-defined and might change for different cases. The semivariogram models were optimized based on minimizing the mean square error, primarily focused on the range parameter. Optimization was performed with the assumption that the model is isotropic and the searching neighbourhood is default with four sectors. Different spatial resolutions were tested, specifically, 5m, 10m, 30m, 50m, 100m, 250m, 500m, 750m, and 1000m to determine if there is a significant drop in accuracy when generating higher spatial resolution images. If there were no significant differences after the interpolation, higher spatial resolution maps can be used to monitor finer areas such as cities or barangays. Then, all the models tested were compared with one another based on their model parameters, R^2, Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). R^2 is a statistical measure that

indicates how well the variance on one variable explains the variance of the other variable RMSE is the standard deviation of prediction errors, while MAPE is the average difference between the estimated and actual values in percent. . The closer the value of R^2 to 1 and the smaller the errors, the more accurate the model estimations are. Lastly, $PM_{2.5}$ and PM_{10} concentration maps were generated using the best models for each spatial resolution.

3. RESULTS AND DISCUSSION

Kriging interpolation techniques, specifically Ordinary Kriging, Simple Kriging, and Universal Kriging were tested with different models: Stable, Circular, Spherical, Exponential, and Gaussian. Fifteen models each for $PM_{2.5}$ and PM_{10} were evaluated using statistical analysis. In general, the $PM_{2.5}$ interpolation models showed R^2 values from 0.76 to 0.92 when compared with the ground monitoring stations as shown in Figure 3. For Ordinary Kriging results for $PM_{2.5}$, spherical semivariogram model showed the highest correlation for interpolated images with spatial resolutions from 5m to 750m resulting in correlation values of 0.68 to 0.83. On the other hand, Stable model and Gaussian model showed the highest correlation value of 0.93 for the 1000m spatial resolution $PM_{2.5}$ image. For the Simple Kriging models, spherical model still showed the highest correlation values of 0.68 to 0.82 with exponential models also showing very high correlation values. For the interpolated $PM_{2.5}$ image with 1000m spatial resolution, the stable model resulted in the highest correlation value of 0.91. For the universal models, correlation values showed similar results of 0.62 to 0.77 across all the models for the different interpolated images. This might be caused by the similar spatial trends calculated across the different models. Overall, for the $PM_{2.5}$ models, 750m interpolated maps showed the lowest correlation. The correlation values seem to increase steadily as the spatial resolution increases from 750m to 10m, followed by a small decrease with the 5m images.

Figure 3. Correlation values for interpolated $PM_{2.5}$ values using Kriging Interpolation with monitoring station data.

For the PM_{10} models, results show correlation values of 0.19 to 0.39 as shown in Figure 4. For the Ordinary Kriging models, Gaussian model displayed the highest correlation values of 0.19 to 0.37 for all the different spatial resolution images, followed by the spherical models. For the Simple Kriging PM_{10} models, Gaussian model only showed the highest value for the 250m to 750m interpolated images with values of 0.24 to 0.39, while Circular model resulted in the best interpolation for the images with spatial resolution less than or equal to 100m with correlation values of 0.24. For the Universal Kriging models, the spherical models showed the highest correlation values across all the interpolated images, resulting in values of 0.10 to 0.12. In general, an increase of correlation value was observed between the 1000m interpolated image and 750m interpolated image. A decrease was seen then from the 500m interpolated image results, while the interpolated images for 5m to 250m showed relatively similar results.

Figure 4. Correlation values for interpolated PM_{10} values using Kriging Interpolation with monitoring station data.

Table 1 and Table 2 show the parameters of the best model for each spatial resolution images for $PM_{2.5}$ and PM_{10}, respectively. RMSE for the $PM_{2.5}$ interpolated models versus monitoring station data resulted in significantly low values of 2.70 to 4.90 $\mu g/m^3$. Moreover, results showed low MAPE values of 7.11 to 15.22%, which is a relatively acceptable indication of accuracy. For the PM_{10} interpolated models, results show relatively low RMSE values from 9.90 to 11.43 $\mu g/m^3$, with MAPE values of 17.41 to 20.14%, which are not relative that larger when compared with the $PM_{2.5}$ results.

Table 1. Best model parameters for the interpolated $PM_{2.5}$ images.

Spatial resolution (m)	Best model	R^2	RMSE ($\mu g/m^3$)	MAPE (%)
5	Ordinary - Spherical	0.8269	4.2209	12.7030
10	Ordinary - Spherical	0.8310	4.1711	12.5618
30	Ordinary - Spherical	0.8315	4.1640	12.5838
50	Ordinary - Spherical	0.8252	4.2409	12.7027
100	Ordinary - Spherical	0.8095	4.4278	13.1405
250	Simple - Spherical	0.7853	4.7008	13.8264
500	Simple - Exponential	0.7674	4.8931	15.2218
750	Universal - Circular, Spherical, Exponential, Gaussian	0.7700	4.8658	14.5947
1000	Ordinary - Stable, Gaussian	0.9292	2.6992	7.1147

Table 2. Best model parameters for the interpolated PM_{10} images.

Spatial resolution (m)	Best model	R^2	RMSE ($\mu g/m^3$)	MAPE (%)
5	Simple - Circular	0.2476	11.0416	19.4254
10	Simple - Circular	0.2462	11.0524	19.4359
30	Simple - Circular	0.2467	11.0487	19.4308
50	Simple - Circular	0.2399	11.0984	19.5297
100	Simple - Circular	0.2482	11.0373	19.3328
250	Simple - Gaussian	0.2411	11.0897	19.3588
500	Ordinary - Gaussian	0.2853	10.7619	19.0243
750	Simple - Gaussian	0.3945	9.9053	17.4064
1000	Ordinary - Gaussian	0.1935	11.4321	20.1408

Figure 5 and Figure 6 display the scatter plots of the best models for $PM_{2.5}$ and PM_{10}, respectively, where the monitoring station data is represented by the y-axis and the interpolated PM values by the x-axis. Scatter plots, especially for the 5m to 500m images, showed no significant difference and displayed similar clustering. Values for $PM_{2.5}$ are concentrated in the 20 to 40 $\mu g/m^3$ range. On the other hand, PM_{10} scatter plots showed a more distributed clustering and variance from the regression line. PM_{10} values concentrated in the 30 to 50 $\mu g/m^3$ range, but implied outlier points at the 20 and 60 $\mu g/m^3$ value of the monitoring station data.

Figure 7 and 8 show the interpolated $PM_{2.5}$ and PM_{10} concentration maps with the highest spatial resolution of 5m. High concentrations of PM are shown in bright red to white while low concentration values are shown in dark red-violet to black. Interpolated maps showed relatively similar results across all the different models of each Kriging type. Universal Kriging maps resulted to smoother and less sharp images than the Ordinary and Simple kriging maps. Maps showed high concentrations in the Manila area and Navotas fishport in general. Moreover, high concentration values can also be found near Quezon Memorial Circle, and along SkyWay and National Road 1. In general, maps show a smooth texture, which is a result of interpolating pixels to smaller pixel sizes. Rough and sharp patches can be also seen at the edges of the maps, especially in the Simple Kriging maps of $PM_{2.5}$ concentration, as those areas are water bodies and have no data points.

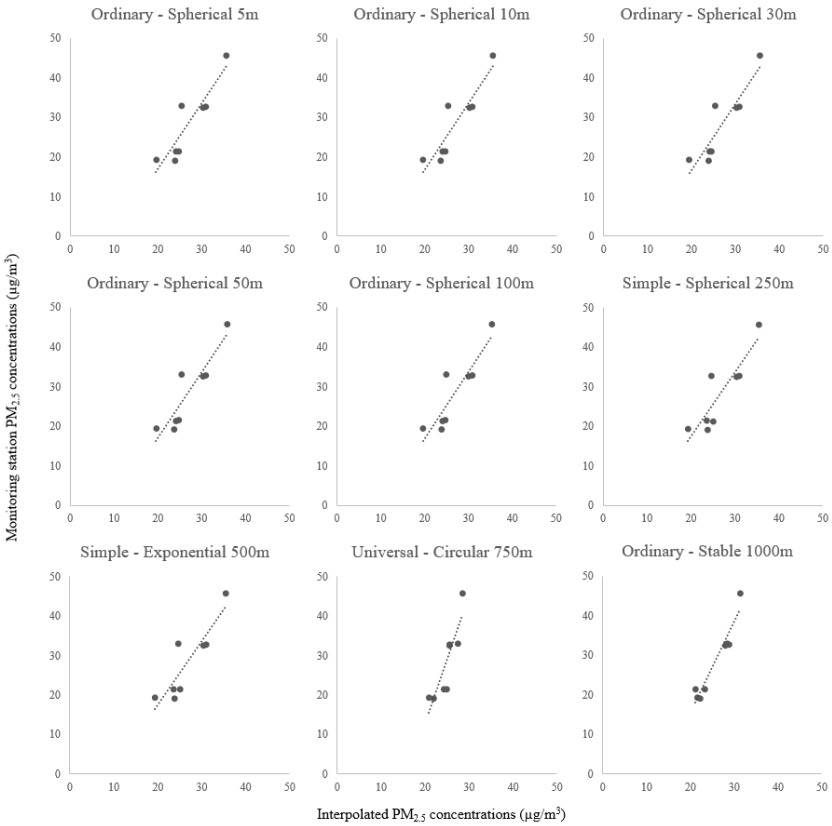

Figure 5. Monitoring Station $PM_{2.5}$ (μg/m^3) versus Interpolated $PM_{2.5}$ (μg/m^3) of the best models for each spatial resolution images.

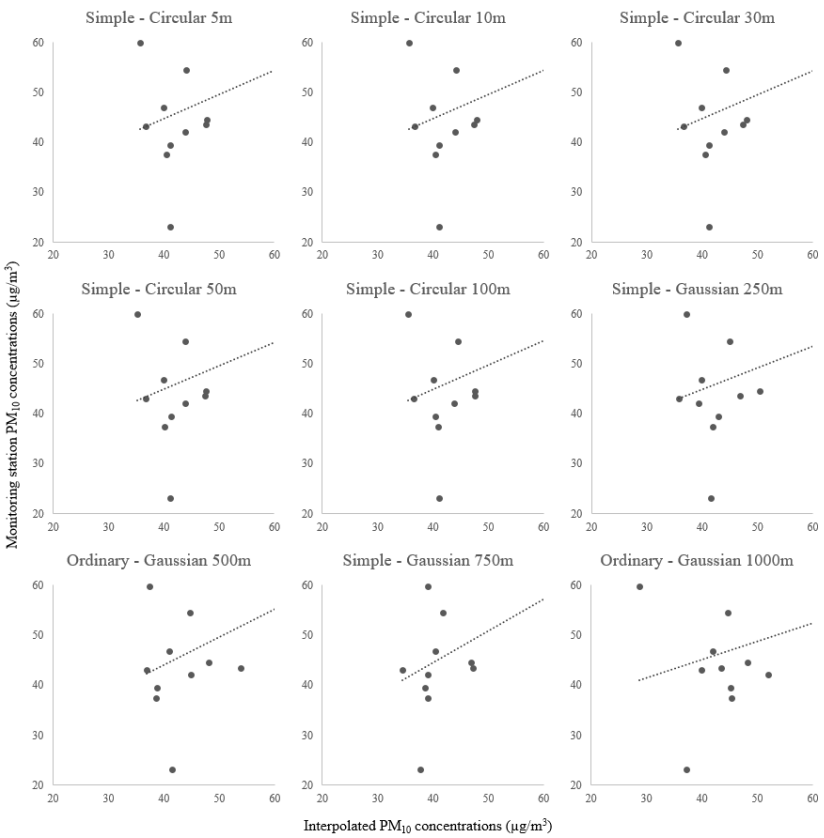

Figure 6. Monitoring Station PM_{10} (μg/m^3) versus Interpolated PM_{10} (μg/m^3) of the best models for each spatial resolution images.

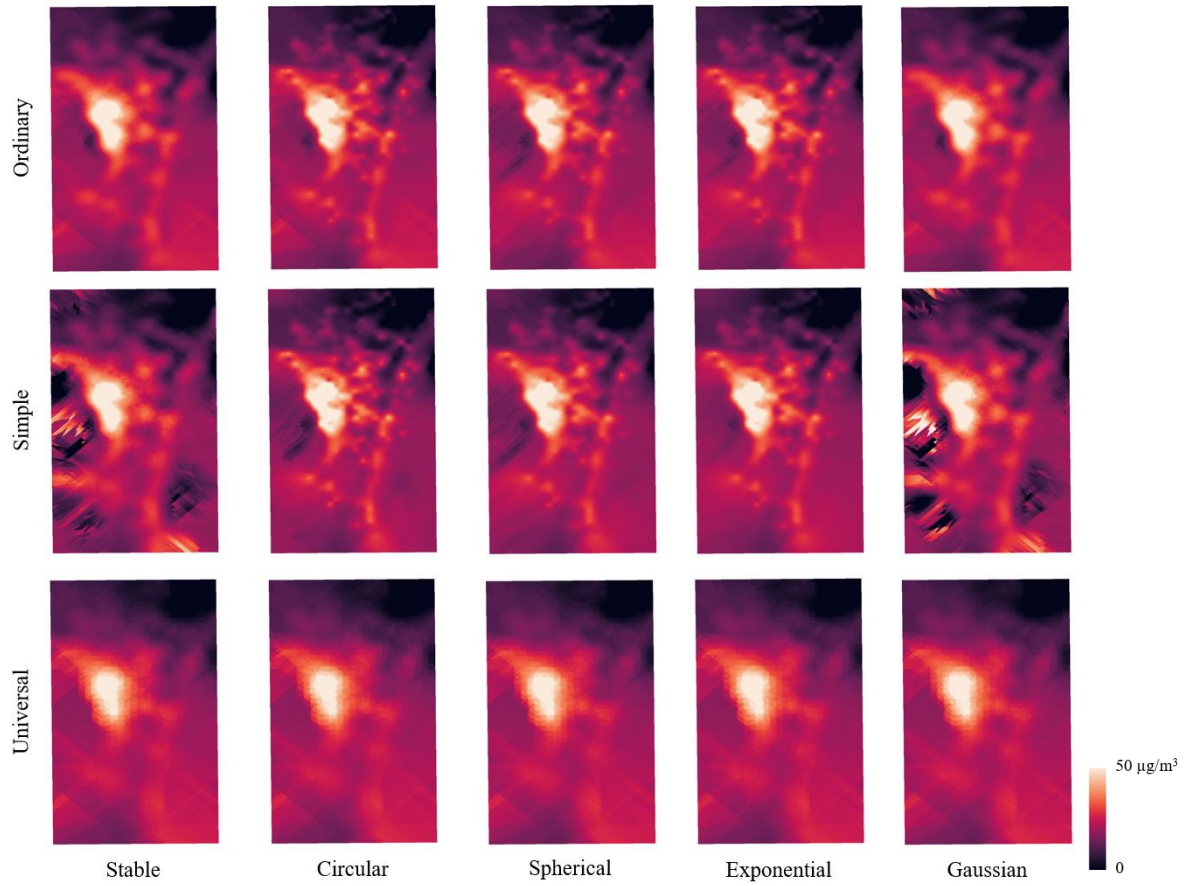

Figure 7. Interpolated $PM_{2.5}$ maps with 5-meter spatial resolution using Kriging Interpolation.

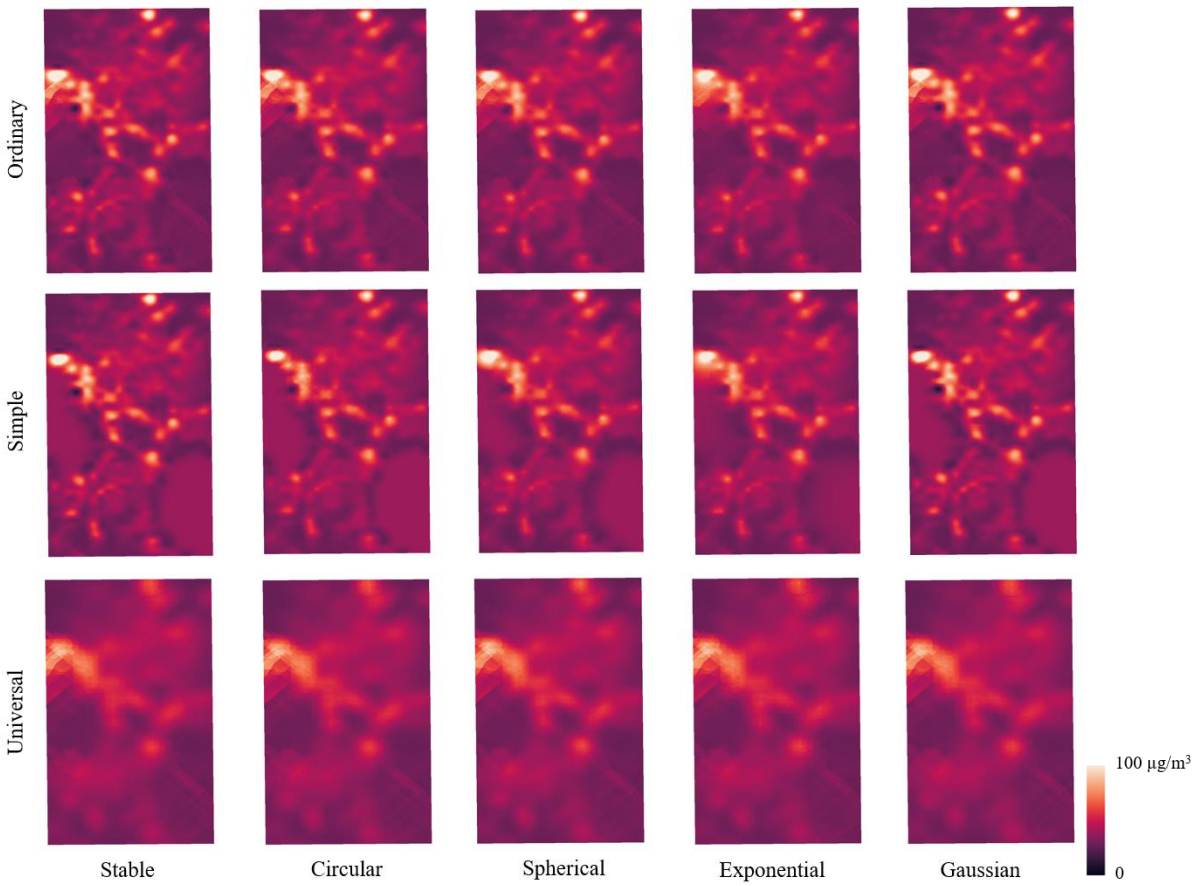

Figure 8. Interpolated PM_{10} maps with 5-meter spatial resolution using Kriging Interpolation.

4. CONCLUSIONS

Higher spatial resolution maps were generated using different kriging interpolation techniques and models. $PM_{2.5}$ models showed R^2 values from 0.76 to 0.92 while PM_{10} models resulted to R^2 values from 0.19 to 0.39. Ordinary – Spherical model showed the best results for interpolated maps with resolutions of 5m to 100m for $PM_{2.5}$ while Simple – Circular model for PM_{10}. For resolutions of 100m to 1000m, Simple and Ordinary kriging interpolation with Gaussian model showed the best results for PM_{10} interpolation. Calculate RMSE and MAPE of the models also showed relatively low values indicating acceptable models. In general, the various Kriging Interpolation types and models showed relatively similar results for the PM images with different spatial resolutions. This indicates that the data itself have high spatial autocorrelation, and even though some of the models showed higher accuracy than the other model, Kriging Interpolation in general is useful in estimating PM concentrations in higher spatial resolution, especially in the case of $PM_{2.5}$. The generated higher spatial resolution images can be used to compliment ground monitoring station data in detecting areas with high concentrations in finer details such as city or barangay inspections. Researchers can explore other modeling types and parameter optimization to improve the study. Longer periods of observation can also be analysed to determine further the effectiveness of the Kriging Interpolation models.

5. ACKNOWLEDGEMENT

This research was done as part of the research project Ambient Air Remote Sensing, Modeling, and Visualization Environment (AiRMoVE) implemented by the University of the Philippines Training Center for Applied Geodesy and Photogrammetry (UP TCAGP), through the support and funding of the Department of Science and Technology (DOST) of the Republic of the Philippines and the Philippine Council for Industry, Energy, and Emerging Research and Development (PCIEERD).

6. REFERENCES

Abdel-Sattar, A., 2019. Monitoring Air Pollution Using Satellite Data. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 772–780.

Bagtasa, G., & Yuan, C.-S., 2020. Influence of local meteorology on the chemical characteristics of fine particulates in Metropolitan Manila in the Philippines. Atmospheric Pollution Research, 11(8), pp. 1359–1369. https://doi.org/10.1016/j.apr.2020.05.013

Gong, G., Mattevada, S., & O'Bryant, S. E., 2014. Comparison of the accuracy of Kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. Environmental Research, 130, pp. 59–69. https://doi.org/10.1016/j.envres.2013.12.005

GISGeography, 2023. *Inverse distance weighting (IDW) interpolation*. GIS Geography. https://gisgeography.com/inverse-distance-weighting-idw-interpolation/

Kamboj, K., Sisodiya, S., Mathur, A. K., Zare, A., & Verma, P., 2022. Assessment and spatial distribution mapping of criteria pollutants. *Water, Air, & Soil Pollution*, 233(3). https://doi.org/10.1007/s11270-022-05522-y

Li, L., & Revesz, P., 2004. Interpolation methods for spatio-temporal geographic data. *Computers, Environment and Urban Systems*, 28(3), pp. 201–227. https://doi.org/10.1016/s0198-9715(03)00018-8

Mabit, L., & Bernard, C., 2007. Assessment of spatial distribution of fallout radionuclides through Geostatistics Concept. *Journal of Environmental Radioactivity*, 97(2–3), pp. 206–219. https://doi.org/10.1016/j.jenvrad.2007.05.008

McGee, D.W., 1991. *The application of Statistical Kriging to improve satellite imagery resolution* (thesis). Defense Technical Information Center, Ft. Belvoir.

Milillo, T. M., Sinha, G., & Gardella, J. A., 2012. Use of geostatistics for remediation planning to transcend urban political boundaries. *Environmental Pollution*, 170, pp. 52–62. https://doi.org/10.1016/j.envpol.2012.06.006

Myllyvirta, L., Thieriot, H., & Suarez, I., 2023. Estimating the Health & Economic Cost of Air Pollution in the Philippines. Independent Report: PH Econ Cost, Centre for Research on Energy and Clean Air.

Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA). (2014). *Climate Types*. Climate of the Philippines. https://www.pagasa.dost.gov.ph/information/climate-philippines

Shamo, B., Asa, E., & Membah, J., 2015. Linear spatial interpolation and analysis of annual average daily traffic data. *Journal of Computing in Civil Engineering*, 29(1). https://doi.org/10.1061/(asce)cp.1943-5487.0000281

Tolentino, P. L., Poortinga, A., Kanamaru, H., Keesstra, S., Maroulis, J., David, C. P., & Ritsema, C. J. (2016). Projected impact of climate change on hydrological regimes in the Philippines. *PLOS ONE*, 11(10). https://doi.org/10.1371/journal.pone.0163941

Wong, D. W., Yuan, L., & Perlin, S. A., 2004. Comparison of spatial interpolation methods for the estimation of Air Quality Data. *Journal of Exposure Science & Environmental Epidemiology*, 14(5), pp. 404–415. https://doi.org/10.1038/sj.jea.7500338

Zhou, Y., & Michalak, A. M., 2009. Characterizing attribute distributions in water sediments by geostatistical downscaling. *Environmental Science & Technology*, 43(24), pp. 9267–9273. https://doi.org/10.1021/es901431y