Spatial interpolation of PM2.5 contaminant using Ordinary Kriging and Support Vector Machine

Felipe Pastén[1], Constanza Irarrázabal[1], Carola Blazquez[2], and Raquel Jiménez[1]

[1]Universidad Andrés Bello, Santiago, Chile, {f.pastncceres, c.irarrzabalpuelle}@uandresbello.edu, rjimenez@unab.cl
[2]Universidad Andrés Bello, Viña del Mar, Chile, cblazquez@unab.cl

***Abstract*—Exposure to air pollution such as particulate matter less than 2.5 micrometers (PM2.5) can produce different types of disease. This study uses mobile measurements of PM2.5 contaminant due to wood burning during winter nights in the conurbation of Temuco and Padre Las Casas in southern Chile. The geostatistical tool *Ordinary Kriging* (OK) and machine learning *Support Vector Machine* (SVM) are employed to estimate an interpolated surface of PM2.5 in this conurbation. Overall, the results using OK indicate spatial variability of PM2.5 concentrations in the conurbation with high values toward the west and east areas of Temuco and some smaller areas of Padre Las Casas. However, the results of spatial interpolation with SVM vary depending on the method used to select the covariates. The best covariate selection for the SVM includes variables related to residential density and local roads within different buffer sizes. Cross-validation analysis suggests that OK outperforms the SVM algorithm when estimating the PM2.5 surface. In addition, the aforementioned results vary depending on the level of aggregation of the data. The results from this study may be used by authorities to implement environmental actions in areas with high PM2.5 concentrations, and properly allocate resources to reduce air pollution in these areas. Future research should include the implementation of other types of machine learning techniques and the use of additional variables that may impact the generation of PM2.5 from wood burning.**

Keywords: Air pollution, Ordinary Kriging, Support Vector Machine, Spatial interpolation

I. Introduction

Ambient air pollution is one of the main causes of health problems worldwide. According to the World Health Organization, approximately 7 million people die every year from exposure to air contaminants, specifically, particulate matter less than 2.5 micrometers (PM2.5) [1]. Short- and long-term exposure to this contaminant may lead to different types of diseases such as stroke, pulmonary disease, lung cancer, asthma, and respiratory infections [2].

Chile is one of the most contaminated countries in Latin America with high PM2.5 concentrations. Seven out of 20 most contaminated cities in the region are Chilean. These cities have high levels of air pollution, particularly in the south of Chile due to residential wood burning that is used for cooking and heating in the winter season.

The conurbation of Temuco and Padre Las Casas is located in southern Chile and has a population of 410,520 inhabitants and an area of 864.7 km^2 [3]. In this conurbation, approximately 90% of the PM2.5 emissions are attributed to the use of household wood burning. In 2015, this conurbation was declared a saturated zone for PM2.5, reaching an average level of this pollutant of 400 $\mu g/m^3$ [4]. According to the Chilean air quality regulations, this level of PM2.5 exceeds 170 $\mu g/m^3$, which is classified as an environmental emergency. As a result of this significant exposure to PM2.5 contaminant, the population in this conurbation suffer from acute respiratory infections and other health problems [5] [6].

Currently, an air quality monitoring network that consists of three stationary monitoring stations is employed to collect daily PM2.5 data in Temuco and Padre Las Casas. However, stationary systems do not capture intra-urban spatial variability in urban areas (i.e., areas with high and low concentrations of PM2.5). Thus, mobile measurements of the PM2.5 contaminant captured by pedestrians, cyclists, and drivers have been used to characterize the spatial distribution of PM2.5 concentrations and population exposure. In this study, we use PM2.5 data that were collected using a vehicle in a mobile measurement campaign conducted in the conurbation of Temuco and Padre Las Casas during winter nights in 2016. Refer to the work in [7] for more details on this campaign.

The aforementioned mobile measurements do not present a complete spatial coverage of the conurbation (i.e., there are locations with no data), and thus, it is not suffice for generating a surface with the spatial distribution of PM2.5. Traditional interpolation methods and machine learning techniques have been used for this purpose [8] [9]. Therefore, the objective of this study is to perform a spatial interpolation of the PM2.5 contaminant using the collected data from the mobile measurement campaign in Temuco and Padre Las Casas. In order to do so, the geostatistical tool *Ordinary Kriging* (OK) and machine learning *Support Vector Machine* (SVM) were implemented using the Smart-Map plug-in in QGIS software [8]. Although this plug-in has been applied to different research areas (e.g. prediction of soil chemical attributes [8], solid organic carbon [10], and noise pollution map [11]), we are not aware of any study that has used the Smart-Map plug-in to estimate PM2.5 concentrations in urban areas.

II. Literature Review

OK is a common geostatistical tool employed for predicting PM2.5 concentrations. For example, [12] used OK to estimate PM2.5 concentrations in Surabaya, Indonesia, and reported high accuracy in the PM2.5 prediction results. In another study, [9] proposed the inclusion of wind direction in the interpolation algorithm OK, and concluded that this inclusion presents more stable and more accurate results than when using OK alone. When comparing OK to other interpolation methods with air pollution data from Tehran, [13] concluded that OK presents less errors than other techniques such as inverse distance weighting for estimating spatial variations of PM2.5.

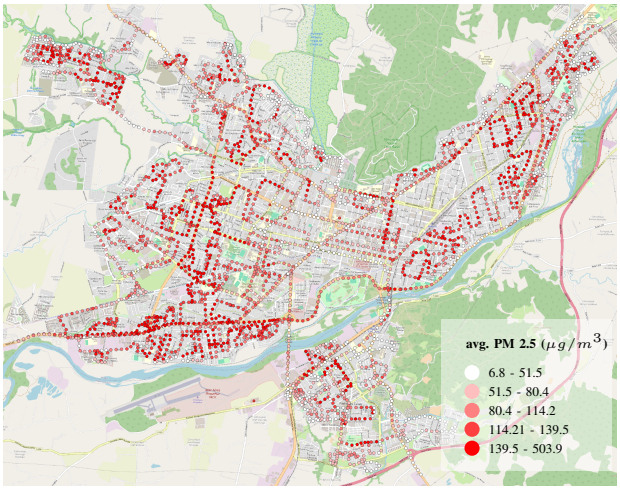
Among different machine learning algorithms, studies have used SVM to estimate PM2.5 concentrations. For example, [14] evaluated SVM to forecast ground-level PM2.5 in Bogotá, Colombia, and their results suggested that SVM describes a number of complex relationships between topography, PM2.5 contaminant, and several meteorological covariates. In the study by [15], the authors combined genetic algorithm and SVM for estimating PM2.5 concentrations, and concluded that their proposed algorithm outperforms conventional models for PM2.5 predictions. Yet in another study, [16] used SVM to predict metal concentrations in PM2.5 and their results indicated that the main urban areas of Nanjing, China showed that the highest heavy metal concentrations occurred nearby industrial and traffic pollution sources.

Some studies have investigated the spatial variability of the PM2.5 contaminant due to wood burning in Chile. For example, [17] obtained interpolated surfaces of PM2.5 to detect spatial clusters with high values of PM2.5 concentrations in the conurbation of Temuco and Padre Las Casas. In the study by [4], the authors estimated a surface of PM2.5 concentrations in the same conurbation by employing the OK interpolation technique and concluded that Temuco presents higher PM2.5 concentrations than Padre Las Casas. In this study, in addition to OK interpolation method, we also use SVM to predict concentrations of PM2.5 in Temuco y Padre Las Casas. Subsequently, comparison results are presented.

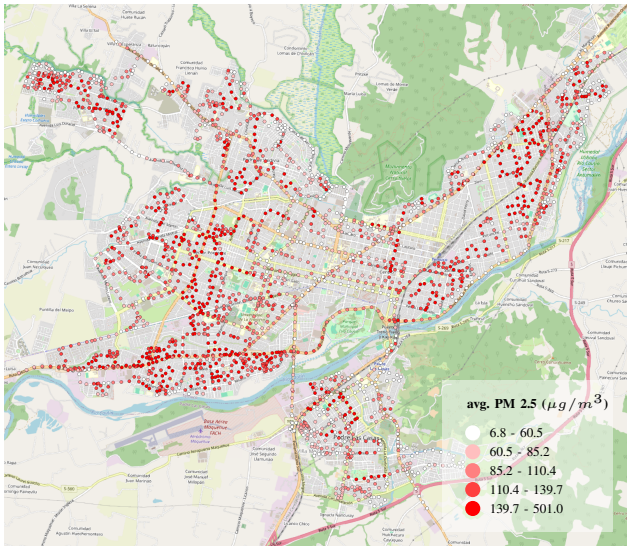
III. Data analysis

As aforementioned, the dataset used in this study comprises PM2.5 measurements collected with mobile sensors following different monitoring routes in the conurbation of Temuco and Padre Las Casas, as described in [7]. First, the data was processed to eliminate measurements with large errors (e.g., measurements located very distant from the conurbation or on the river), and then the measurements were moved (snapped) to the closest road segments. Second, due to the large amount of measurements that was collected in the campaign and to avoid data redundancy, these measurements were aggregated at the 100-meter and 150-meter segment levels along the roadway network, similar to the work by [18]. Third, centroids were obtained for each segment, and average values for snapped PM2.5 measurements were calculated for each centroid.

Fig. 1a) and 1b) present the average PM2.5 measurements at each centroid for the 100-meter and 150-meter segments, respectively. These figures show a total of 3,129 and 2,302 centroids for the 100-meter and 150-meter segments, respectively. PM2.5 contaminant values increase from white to darker red colors. Note that OK and SVM used these average PM2.5 measurements for generating the interpolated surface of the PM2.5 contaminant.



(a) 100-meter segments



(b) 150-meter segments

Fig. 1 Distribution of aggregated PM 2.5 measurements in the conurbation of Temuco and Padre Las Casas

Additionally, an analysis is required to identify suitable variables to be used in the predictive SVM model. Thus, variable values from the groups in Table 1 (Demographic, Transportation, Land Use, Environmental, and Landsat 8 in-

dices) were computed for each centroid. This table shows that the variable values were obtained within three buffer sizes (100, 200, and 300 meters), from surface density such as Kernel Estimation Density, from distances to closest roads or railroads, or pixels from satellite images.

The demographic variables and the roadway network with different road types were obtained from the 2017 Census [3]. Land use information was provided by the 2002 Census [19] and the tax appraisal data from the Chilean Internal Revenue Service [20]. The environmental variables were downloaded from [21]. Finally, the Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) for both annual and winter periods were derived from Landsat 8 satellite data from Google Earth Engine. Pixel values from the satellite images were extracted for every centroid. During this extraction process, some centroids were not able to extract a pixel value due to cloud cover in the winter satellite images. In order to address this issue, a nearest-neighbor Inverse Distance Weighted (IDW) approach was implemented to obtain the missing values for the centroids.

TABLE I
Description of variables used in the analysis

Variable Group	Description
Contaminant	
avg_pm25	Average PM 2.5 concentrations
Demographic	
pop_100, pop_200, pop_300	Population in 100, 200, 300-m buffers
pop_den	Population density
dwe_100, dwe_200, dwe_300	Dwellings in 100, 200, 300-m buffers
dwe_den	Dwelling density
wom_100, wom_200, wom_300	Women in 100, 200, 300-m buffers
wom_den	Women density
men_100, men_200, men_300	Men in 100, 200, 300-m buffers
men_den	Men density
woo_100, woo_200, woo_300	Woodstoves in 100, 200, 300-m buffers
woo_den	Woodstove density
Transportation	
tra_100, tra_200, tra_300	Transit in 100, 200, 300-m buffers
tran_dist	Distance to transit
loc_100, loc_200, loc_300	Local roads in 100, 200, 300-m buffers
loc_dist	Distance to local roads
maj_100, maj_200, maj_300	Major roads in 100, 200, 300-m buffers
maj_dist	Distance to major roads
hwy_100, hwy_200, hwy_300	Highways in 100, 200, 300-m buffers
hwy_dist	Distance to highways
htr_100, htr_200, htr_300	High traffic in 100, 200, 300-m buffers
htr_dist	Distance to high traffic congestion
ltr_100, ltr_200, ltr_300	Low traffic in 100, 200, 300-m buffers
ltr_dist	Distance to low traffic congestion
rai_100, rai_200, rai_300	Railroad in 100, 200, 300-m buffers
rai_dist	Distance to railroad
Land Use	
res_100, res_200, res_300	Residential in 100, 200, 300-m buffers
res_den	Residential density
gre_100, gre_200, gre_300	Green areas in 100, 200, 300-m buffers
gre_den	Green areas density
com_100, com_200, com_300	Commerce in 100, 200, 300-m buffers
com_den	Commerce density
ind_100, ind_200, ind_300	Industry in 100, 200, 300-m buffers
ind_den	Industry density
ser_100, ser_200, ser_300	Services in 100, 200, 300-m buffers
ser_den	Services density
oth_100, oth_200, oth_300	Other land uses in 100, 200, 300-m buffers
oth_den	Other land uses density
tax_100, tax_200, tax_300	Tax appraisal in 100, 200, 300-m buffers
tax_den	Tax appraisal density
Environmental	
alt	Altitude
win	Wind speed
hum	Relative humidity
pre	Precipitation
Landsat 8 indices	
NDVI_winter, NDVI_annual	Normalized Difference Vegetation Index
LST_winter, LST_annual	Land Surface Temperature

IV. Methodology

Both OK and SVM that were used to generate interpolated surfaces of PM2.5 contaminant in the conurbation of Temuco and Padre Las Casas are described in this section. Note that the accuracy in the performance of the results is evaluated using R^2 and RMSE.

The OK technique interpolates values at unmeasured locations based on the measured locations. First, a semivariogram model is selected that best fits the PM2.5 data. This model determines the distance at which the data are no longer autocorrelated (i.e., spatial dependence) [4]. Subsequently, OK is implemented to obtain a continuous PM2.5 surface using the selected semivariogram model. OK is commonly expressed by (1), where $Z^*(x_o)$ is the estimated value at location x_o, λ_i is the weighted coefficient at location i, $Z(x_i)$ are the measured values at location i, and n is the number of measured values [22]. Finally, a cross-validation is executed to assess the accuracy of the estimated interpolated surface with respect to the PM2.5 values at each location.

$$Z^*(x_o) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

SVM is a machine learning technique that is also used in this study for the prediction of PM2.5 measurements. SVM generates a predictive model using a hyperplane that maximizes the separation margin between classes and the risk minimization principle [14].

First, covariates for the SVM method are selected based on the correlation using the Spearman correlation coefficient and the spatial autocorrelation using bivariate Moran's I statistic. While the Spearman correlation coefficient measures the relationship between each covariate and the average PM2.5 contaminant, the bivariate Moran's I index measures the spatial correlation (clustering) between each covariate and the average PM2.5 contaminant. For the Spearman correlation, those covariates that are highly correlated with the average PM2.5 contaminant with values greater than $\left|0.2\right|$ were selected. Similarly, covariates with correlation values greater than 0.2

and 0.3 were selected using the Moran's I statistic to ensure high level of clustering.

Subsequently, the hyperparameters need to be tuned according to the type and variation of the data. In this study, the values of regularization (C) and gamma hyperparameters are optimized based on a systematic grid search method, as in [8]. In addition, the Radial Basis Function Kernel hyperparameter is selected since this function fits most data. Once the model fits the data, the interpolated surface for the PM2.5 pollutant is generated based on IDW of the closest neighbors to the location to be interpolated. Finally, a cross-validation is performed to obtain the accuracy of the interpolated surface of PM2.5.

V. Results

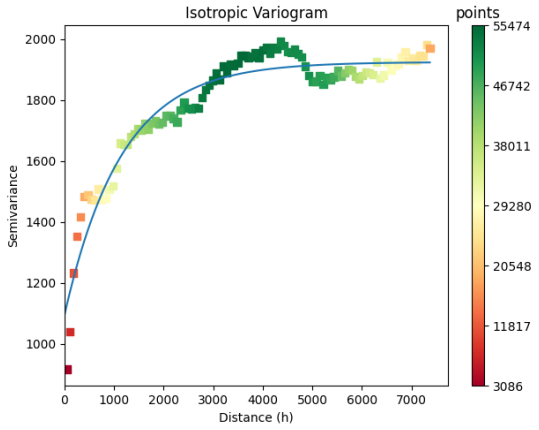
A. Ordinary Kriging

As aforementioned, a semivariogram model needs to be selected that best fits the data. Fig. 2a) and 2b) present the semivariograms for the 100-m and 150-m segments, respectively, indicating that an exponential curve best fits both semivariograms. This curve has a high accuracy with R^2 of 0.928 and 0.933 for 100-m and 150-m segments, respectively.

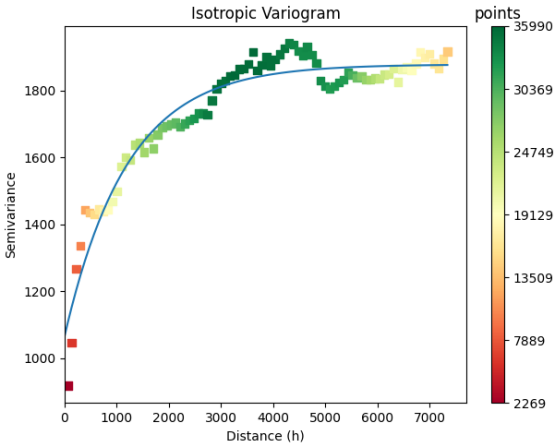
Table II presents the main characteristic values of the semivariogram (i.e., sill, nugget, and range). As mentioned in [4], the sill is the maximum value of the curve, the nugget is the value at which the curve intercepts in the Y axis, and the range is the location at which the measured points stop contributing to the estimation of the unmeasured points.

OK used the values in Table II to obtain the interpolated maps of the PM2.5 contaminant for the 100-m and 150-m segments shown in Fig. 3a) and 3b), respectively. These figures show high spatial variability with similar locations of low and high concentrations of PM2.5. High concentrations are observed toward the east and west sides of Temuco, and low concentrations of PM2.5 are perceived in the center of Temuco. Padre Las Casas present lower spatial variability of PM2.5 than in Temuco, as in [4].

A cross-validation was performed to determine the performance of OK in the estimation of the PM2.5 interpolated surface. The interpolated values of PM2.5 were computed and compared with locations with average PM2.5 at the segment level. Fig. 4a) and 4b) present the relationship between the observed and predicted values of PM2.5 using the OK technique. In this case, R^2 and RMSE were calculated for each interpolated map with 100-m and 150-m segments. Slightly better results were obtained in the estimation of the PM2.5 interpolated surface with the 100-m segments than with the 150-m segments since higher R^2 and lower RMSE are observed in these results.



(a) 100-m segments



(b) 150-m segments

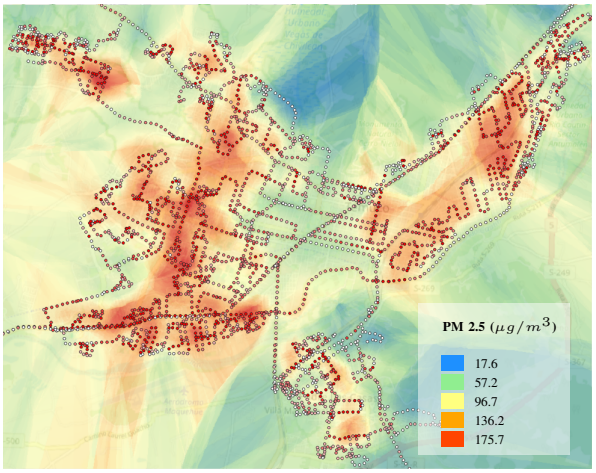
Fig. 2 Semivariogram models for different segment sizes

TABLE II
Results of the semivariogram model

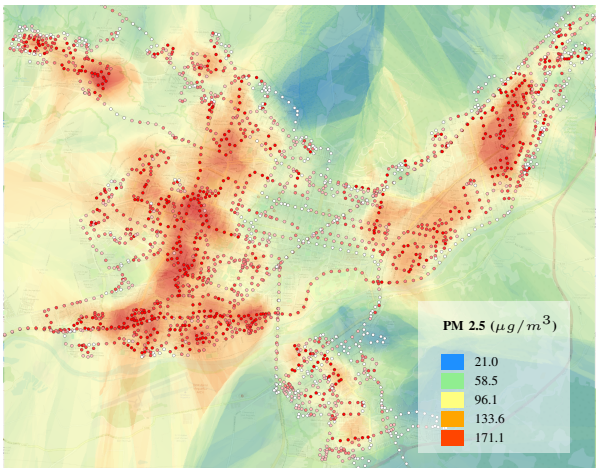
Variable	100-m Segment	150-m Segment
Nugget (m)	1096.09	1066.93
Sill (m)	1925.24	1878.09
Range (m)	3457.81	3605.88

B. Support Vector Machine

As aforementioned, before implementing the SVM technique, the covariates are selected using the Spearman correlation coefficient and the Moran´s I statistic. Based on the correlation analysis, a subset of variables was selected to ensure statistical significance and reduce redundancy within each group shown in Table I.

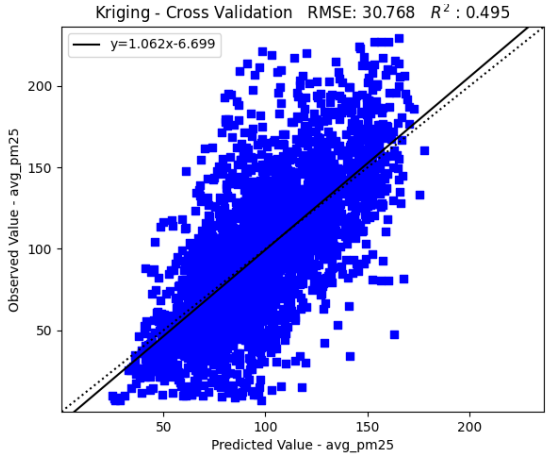


(a) 100-m segments

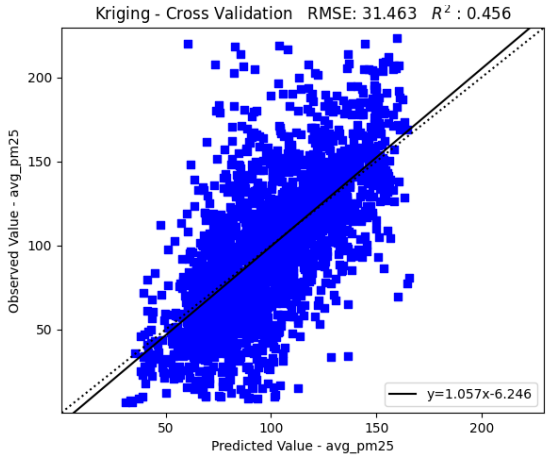


(b) 150-m segments

Fig. 3 Surface interpolation of PM2.5 using OK



(a) 100-m segments



(b) 150-m segments

Fig. 4 Cross-validation of the interpolated surface of PM2.5 using OK

Table III shows the list of covariates that were selected for the SVM using the Spearman correlation analysis for 100-m and 150-m segments. Only covariates with correlation values greater than $|0.2|$ are listed in this table. For 100-m segments, the highest correlation with PM2.5 are the local roads within 300-m buffers and residential density with 0.44 and 0.41, respectively, followed by green area density and population within 300-m buffers. Note that the winter NDVI presents a negative correlation with average PM2.5, suggesting that the amount of PM2.5 decrease with vegetation. The same covariates were selected for 150-m segments with very similar correlation values, except for the covariate related to dwellings in 300-m buffers with a correlation of 0.28 that was incorporated into the list.

TABLE III
Selected covariates for 100-m and 150-m segments using Spearman correlation

Variable	Description	Spearman Correlation
100-m segment		
loc_300	Local roads in 300-m buffer	0.44
res_den	Residential Density	0.41
gre_den	Green Areas Density	0.33
pop_300	Population in 300-m buffer	0.29
NDVI_winter	Winter NDVI	-0.24
150-m segment		
loc_300	Local roads in 300-m buffer	0.44
res_den	Residential Density	0.41
gre_den	Green Areas Density	0.32
pop_300	Population in 300-m buffer	0.29
dwe_300	Dwellings in 300-m buffer	0.28
NDVI_winter	Winter NDVI	-0.24

Note: p-value < 0.05 for all reported results.

TABLE IV
Selected covariates for 100-m and 150-m segments using Moran's I statistic

Variable	Description	Moran's I
100-m segment		
loc_300	Local roads in 300-m buffers	0.394
loc_200	Local roads in 200-m buffers	0.369
res_den	Residential density	0.344
loc_100	Local roads in 100-m buffers	0.326
loc_dist	Distance to local roads	0.250
men_300	Men in 300-m buffers	0.243
pop_300	Population in 300-m buffers	0.240
wom_300	Women in 300-m buffers	0.239
dwe_300	Dwellings in 300-m buffers	0.236
gre_den	Green areas density	0.233
dwe_den	Dwelling density	0.222
men_den	Men density	0.222
wom_den	Women density	0.220
dwe_200	Dwelling in 200-m buffers	0.219
men_200	Men in 200-m buffers	0.217
wom_200	Women in 200-m buffers	0.216
pop_200	Population in 200-m buffers	0.215
pop_den	Population density	0.209
150-m segment		
loc_300	Local roads in 300-m buffers	0.380
loc_200	Local roads in 200-m buffers	0.347
res_den	Residential density	0.330
loc_100	Local roads in 100-m buffers	0.294
men_300	Men in 300-m buffers	0.225
gre_den	Green areas density	0.224
loc_dis	Distance to local roads	0.222
pop_300	Population in 300-m buffers	0.221
wom_300	Women in 300-m buffers	0.221
dwe_300	Dwellings in 300-m buffers	0.217

Note: p-value < 0.05 for all reported results.

TABLE V
Performance results of SVM using Moran's I index and Spearman correlation

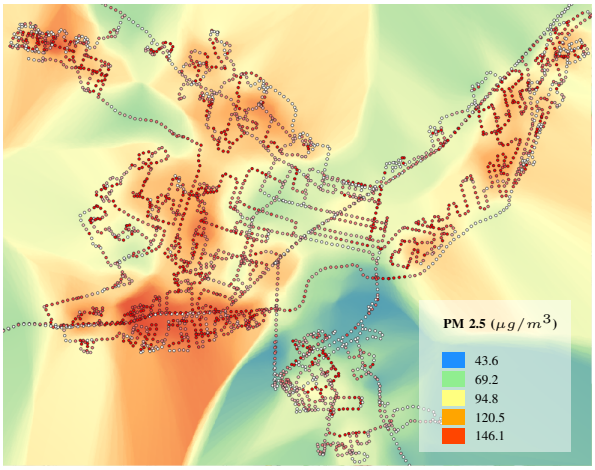
Segment level	Moran's I > 0.2		Moran's I > 0.3		Spearman $> \|0.2\|$	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
100-m segments	38.517	0.207	38.297	0.216	37.744	0.239
150-m segments	37.338	0.232	34.141	0.361	37.313	0.233

In contrast to the Spearman correlation, Table IV indicates that a larger number of variables were selected as covariates using the SVM implementation with the Moran´s I statistic. For both 100-m and 150-m segments, the covariates related to local roads within 100-m, 200-m, and 300-m buffers, and residential density present the same top four highest spatial clustering among all variables. This clustering indicates that locations with high (low) values of PM2.5 are surrounded by covariates with similar high (low) values.

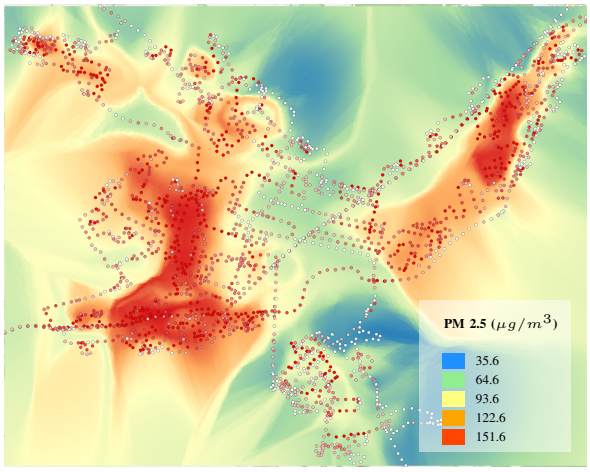
Table V presents the performance results of SVM for 100-m and 150-m segments using Spearman correlation and Moran's I index. This table suggests that improved interpolation results are obtained for 150-m segments when using covariate selected by Moran's I > 0.3 since there is a higher clustering.

Fig. 5 shows the interpolated surface of the PM2.5 contaminant using SVM for the best performance results indicated in Table V, i.e., Spearman correlation for 100-m segments (Fig. 5a) and Moran's I > 0.3 for 150-m segments (Fig. 5b). The interpolated surfaces of PM2.5 shown in these figures display spatial variability in the studied area, as in OK method. Fig. 5a) depicts that the highest estimated PM2.5 values are concentrated in a large area located southwest of Temuco for 100-m segments in an intense red color, and other areas with less intensity. Whereas, Fig. 5b) shows high PM2.5 concentrations scattered in different areas of the conurbation. Note that low PM2.5 values are identified in Padre Las Casas for 100-m segments.

Similar to the OK method, a cross-validation was conducted to determine the performance of SVM when estimating the PM2.5 surface. The association between the observed and predicted values of PM2.5 using SVM are shown in presented in Fig. 6. Fig. 6a) and 6b) also show the values of R^2 and RMSE for the interpolated surfaces of PM2.5 using SVM for 100-m and 150-m segments, respectively. These results suggests that higher R^2 and lower RMSE are obtained for 150-segments with the Moran's I index.

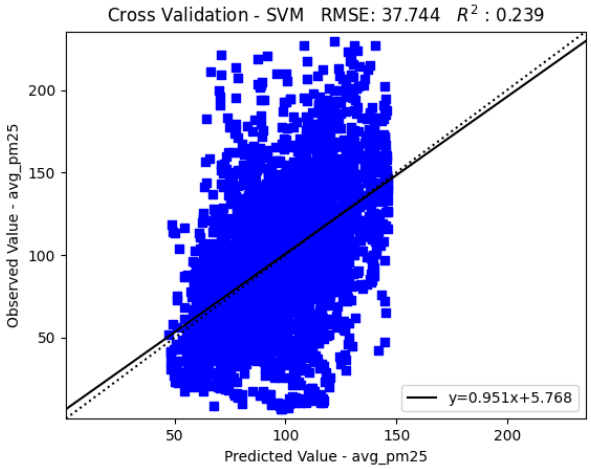


(a) 100-m segments - Spearman

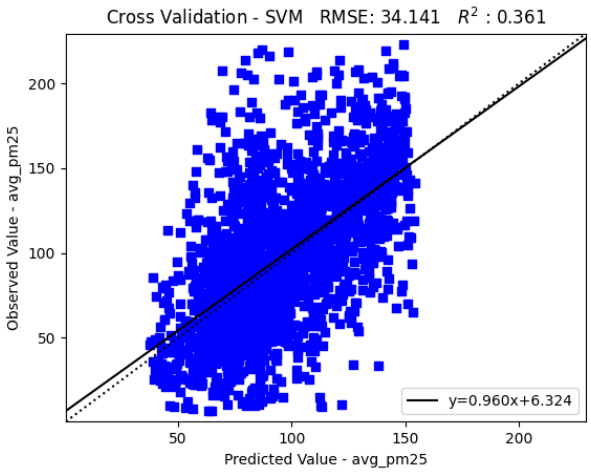


(b) 150-m segments - Moran's I > 0.3

Fig. 5 Interpolated surface of PM2.5 using SVM for the best selection results



(a) 100-m segments - Spearman



(b) 150-m segments - Moran's I > 0.3

Fig. 6 Cross-validation of the interpolated surface of PM2.5 using SVM for the best selection results

VI. Conclusions

This study used Ordinary Kriging (OK) and Support Vector Machine (SVM) to estimate interpolated surfaces of the PM2.5 contaminant. This contaminant is generated from wood burning during winter nights in the conurbation of Temuco and Padre Las Casas, located in southern Chile.

Overall, the results suggest that the outcome of the interpolated surface may vary according to the aggregation of the variables at the segment level along the roadway network. For example, the accuracy of the PM2.5 interpolated surface for data aggregated using 100-m segments is slightly better than using 150-m segments for the OK technique (R^2 = 0.495 and R^2 = 0.456, respectively).

In the selection of the covariates for the SVM algorithm, Spearman correlation and spatial autocorrelation using Moran's I > 0.2 present similar accuracy results (R^2 = 0.232 versus R^2 = 0.233), particularly, for 150-m segments. When comparing the data aggregation level, the best performance results using SVM are obtained when selecting the covariates with Spearman correlation for 100-m segments, and with Moran's I > 0.3 for 150-m segments. Finally, the cross-validation analysis suggests that OK presents lower error in generating the PM2.5 interpolated surface than the SVM technique.

These results may help authorities and policymakers to implement environmental actions to reduce air pollution in zones with high concentrations of PM2.5 contaminant in the studied conurbation. Future research should include the use of additional variables that may impact the generation of PM2.5 in the SVM models. Additional machine learning techniques should be implemented to estimate the spatial interpolation of PM2.5 contaminant and also to compare with OK.

References

[1] World Health Organization, "Noncommunicable diseases," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

[2] European Environment Agency, "How air pollution affects our health," 2024. [Online]. Available: https://www.eea.europa.eu/en/topics/in-depth/air-pollution/eow-it-affects-our-health

[3] Instituto Nacional de Estadística, "Censos de Población y Vivienda: Región de la Araucanía," 2017. [Online]. Available: https://regiones.ine.gob.cl/araucania/estadisticas-regionales/sociales/censos-de-poblacion-y-vivienda

[4] J. C. Cubillos and C. Blazquez, "Predicting PM2.5 in Temuco and Padre Las Casas, Chile using Ordinary Kriging," in *19th LACCEI International Multi-Conference for Engineering, Education Caribbean Conference for Engineering and Technology:" Prospective and Trends in Technology and Skills for Sustainable Social Development" and" Leveraging Emerging Technologies to Construct the Future", LACCEI 2021*. Latin American and Caribbean Consortium of Engineering Institutions, 2021.

[5] B. Á. Escobar, P. C. Farina, J. Navarro-Riffo, C. M. Muñoz, and Á. B. Gaspar, "Comportamientos de autoprotección frente a la contaminación del aire y factores psicosociales, Temuco, Chile," *Revista Internacional de Contaminación Ambiental*, vol. 38, 2022.

[6] C. Molina, R. Toro A, R. G. Morales S, C. Manzano, and M. A. Leiva-Guzmán, "Particulate matter in urban areas of south-central Chile exceeds air quality standards," *Air Quality, Atmosphere & Health*, vol. 10, pp. 653–667, 2017.

[7] E. Blanco, F. Rubilar, M. E. Quinteros, K. Cayupi, S. Ayala, S. Lu, R. Jimenez, J. P. Cardenas, C. Blazquez, J. M. Delgado-Saborit, R. Harrison, and P. Ruiz-Rudolph, "Spatial distribution of particulate matter on winter nights in Temuco, Chile: Studying the impact of residential wood-burning using mobile monitoring," *Atmospheric Environment*, vol. 286, no. 119255, 2022.

[8] G. W. Pereira, D. S. M. Valente, D. M. d. Queiroz, A. L. d. F. Coelho, M. M. Costa, and T. Grift, "Smart-map: An open-source QGIS plugin for digital mapping using machine learning techniques and ordinary kriging," *Agronomy*, vol. 12, no. 6, p. 1350, 2022.

[9] H. Zhang, Y. Zhan, J. Li, C.-Y. Chao, Q. Liu, C. Wang, S. Jia, L. Ma, and P. Biswas, "Using Kriging incorporated with wind direction to investigate ground-level PM2. 5 concentration," *Science of the Total Environment*, vol. 751, p. 141813, 2021.

[10] C. A. Aguirre-Salado, J. Perez-Nieto, C. A. Aguirre-Salado, and A. I. Monterroso-Rivas, "Factors regarding the spatial variability of soil organic carbon in a Mexican small watershed," *Revista de la Facultad de Agronomía de la Universidad de Zulia*, vol. 41, no. 1, 2024.
[11] Z. M. Nasser, A. H. Abedali, and H. A. Alkanaani, "Reliability of smart noise pollution map," *Noise Mapping*, vol. 10, no. 1, p. 20220167, 2023.
[12] D. W. Fitri, N. Afifah, S. Anggarani, and N. Chamidah, "Prediction concentration of PM2. 5 in Surabaya using ordinary Kriging method," in *AIP Conference Proceedings*, vol. 2329, no. 1. AIP Publishing, 2021.
[13] K. Masroor, F. Fanaei, S. Yousefi, M. Raeesi, H. Abbaslou, A. Shahsavani, and M. Hadei, "Spatial modelling of PM2. 5 concentrations in Tehran using Kriging and inverse distance weighting (IDW) methods," *Journal of Air Pollution and Health*, vol. 5, no. 2, pp. 89–96, 2020.
[14] C. Mogollón-Sotelo, A. Casallas, S. Vidal, N. Celis, C. Ferro, and L. Belalcazar, "A support vector machine model to forecast ground-level PM 2.5 in a highly populated city with a complex terrain," *Air Quality, Atmosphere & Health*, vol. 14, pp. 399–409, 2021.
[15] P. Zhang, W. Ma, F. Wen, L. Liu, L. Yang, J. Song, N. Wang, and Q. Liu, "Estimating PM2. 5 concentration using the machine learning GA-SVM method to improve the land use regression model in Shaanxi, China," *Ecotoxicology and Environmental Safety*, vol. 225, p. 112772, 2021.
[16] X. Qian, M. Yang, C. Wang, H. Li, J. Wang *et al.*, "Leaf magnetic properties as a method for predicting heavy metal concentrations in PM2. 5 using support vector machine: A case study in Nanjing, China," *Environmental Pollution*, vol. 242, pp. 922–930, 2018.
[17] M. E. Quinteros, E. Blanco, J. Sanabria, F. Rosas-Diaz, C. A. Blazquez, S. Ayala, J. P. Cárdenas-R, E. A. Stone, K. Sybesma, J. M. Delgado-Saborit *et al.*, "Spatio-temporal distribution of particulate matter and wood-smoke tracers in Temuco, Chile: A city heavily impacted by residential wood-burning," *Atmospheric Environment*, vol. 294, p. 119529, 2023.
[18] M. Abarca, C. Blazquez, and K. Sauer-Brand, "Estimación de los factores que influyen en la generación del contaminante PM2. 5 en Temuco y Padre Las Casas, Chile," in *Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology: Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Society 5.0., LACCEI 2024*. Latin American and Caribbean Consortium of Engineering Institutions, 2024.
[19] Instituto Nacional de Estadística, "Censos de Población y Vivienda," 2002. [Online]. Available: https://www.ine.gob.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda/censo-de-poblacion-y-vivienda
[20] Servicio de Impuestos Internos, "Cartografía Digital SII Mapas," 2020. [Online]. Available: https://www4.sii.cl/mapasui/internet//contenido/index.html
[21] Dirección Meteorológica de Chile, "Explora Capas," 2016. [Online]. Available: https://geonode.meteochile.gob.cl/
[22] M. A. Oliver, R. Webster *et al.*, "Basic steps in geostatistics: the variogram and kriging," Springer, Tech. Rep., 2015.