Predicting PM2.5 in Temuco and Padre Las Casas, Chile using Ordinary Kriging

Juan Carlos Cubillos, BSc [1], and Carola Blazquez, Ph.D.[1]
[1]Universidad Andres Bello, Chile, j.cubillosletelier@uandresbello.edu, cblazquez@unab.cl

Abstract– Ambient air pollution causes serious human health problems. Geostatistical interpolation methods have proven to efficiently estimate exposure to air pollutants. In this study, we employed Ordinary Kriging to predict PM2.5 concentrations due to woodsmoke in the conurbation of Temuco and Padre Las Casas, Chile using mobile measurements conducted during June and July of 2016. Overall, the results suggest that higher PM2.5 concentrations were estimated in Temuco. Spatial differences in high PM2.5 concentrations are observed when examining the PM2.5 estimates by month. The results of this study may help authorities and policymakers to implement environmental actions to reduce air pollution in the studied conurbation.

Keywords-- PM2.5 concentrations, mobile measurements, Ordinary Kriging, Chile

Digital Object Identifier: http://dx.doi.org/10.18687/LACCEI2021.1.1.263
ISBN: 978-958-52071-8-9 **ISSN**: 2414-6390
DO NOT REMOVE

Predicting PM2.5 in Temuco and Padre Las Casas, Chile using Ordinary Kriging

Juan Carlos Cubillos, BSc [1], and Carola Blazquez, Ph.D.[1]
[1]Universidad Andres Bello, Chile, j.cubillosletelier@uandresbello.edu, cblazquez@unab.cl

Abstract*– Ambient air pollution causes serious human health problems. Geostatistical interpolation methods have proven to efficiently estimate exposure to air pollutants. In this study, we employed Ordinary Kriging to predict PM2.5 concentrations due to woodsmoke in the conurbation of Temuco and Padre Las Casas, Chile using mobile measurements conducted during June and July of 2016. Overall, the results suggest that higher PM2.5 concentrations were estimated in Temuco. Spatial differences in high PM2.5 concentrations are observed when examining the PM2.5 estimates by month. The results of this study may help authorities and policymakers to implement environmental actions to reduce air pollution in the studied conurbation.

Keywords-- PM2.5 concentrations, mobile measurements, Ordinary Kriging, Chile

I. Introduction

Particulate matter less than 2.5 micrometers in diameter (PM2.5) is an ambient air pollutant that kills more than 4.2 million people every year around the world, as stated by the World Health Organization [1, 2]. PM2.5 consists of organic substances, dust, metals, chemicals, and soot, which enters the respiratory system and causes different respiratory diseases or cancer, asthma, cardiac problems, allergies, and even premature death [3].

In southern Chile, residential wood combustion from cook stoves and heaters is one of the main sources of the emitted PM2.5 into the air [4]. In particular, an urgent and critical air pollution problem exists in the conurbation of Temuco and Padre Las Casas (PLC) that is located in the Region of Araucanía and has a total population of 358,541 inhabitants [5]. More than 90% of the households in this conurbation have woodstoves that yield large amounts of PM2.5 contaminant, and thus, in 2015, was declared saturated zone for PM2.5. The wood combustion of diesel and gasoline or oil is commonly employed in this conurbation, reaching high average PM2.5 concentration levels of over 400 $\mu g/m^3$ [6]. As a result, a large share of population in Temuco and PLC suffers from acute respiratory infections [7, 8].

A mobile monitoring campaign of PM2.5 was performed during the winter of 2016 to describe and characterize the spatial distribution of this air pollutant when woodburning is mostly employed in the conurbation of Temuco and PLC. Using the measurements from this campaign, we studied the spatial variability of PM2.5 in this conurbation with a geostatistical tool named Ordinary Kriging. This tool is the most-used kriging method and is useful for predicting PM2.5 concentrations in those geographic locations without any measurements of monitoring data using measured values from neighboring locations [9]. This method has proven a higher confidence level in the results since it has a lower error range [10].

The results of this study will help authorities and policymakers in the decision-making process on early warnings and prevention of air pollution, particularly in those areas of the conurbation that have the highest PM2.5 concentration levels.

II. Literature Review

Different models have been used to predict PM2.5 concentrations such as land use regression [11, 12], machine learning [13, 14], and geostatistical interpolation methods [9, 15, 16, 17, 18]. In particular, Kriging is a geostatistical method that has been widely employed to generate a surface of PM2.5 with several measurement points. For example, reference [9] used Ordinary Kriging as a space-time interpolation method of PM2.5 concentrations to comprehend the behavior of this pollutant in Bogotá, Colombia. Reference [15] employed Kriging and inverse distance weighting (IDW) to generate a surface of PM2.5 contaminant using fixed monitoring stations also in Bogotá, Colombia. The authors identified zones with high levels of PM2.5 concentrations and reported that Kriging presented more satisfactory results compared to IDW. Similarly, reference [16] used Kriging and IDW to predict PM2.5 at unmonitored locations and to visualize spatial and temporal variability of this contaminant. [17] studied the concentrations of PM2.5 spatially and temporally in Santiago, Chile. The authors identified high PM2.5 pollutant in the west part of the city, where population density is higher and income levels are lower. In another study, reference [18] implemented Kriging with an external drift to predict PM2.5 concentrations in Beijing, China. Finally, reference [19] employed Ordinary Kriging interpolation and other statistical methods to explore spatiotemporal variations of PM2.5 concentrations in Beijing, China.

Although authors have studied the air pollution in Temuco [20, 21, 22], none of these authors have employed geostatistical interpolation methods such as Kriging. We are not aware of any study that employs Ordinary Kriging to identify the variability of PM2.5 concentrations over space and time in the conurbation of Temuco and PLC.

III. Data Description and Preparation

This study employed PM2.5 measurements collected during a mobile sampling campaign in June and July of 2016 in Temuco and PLC. Since woodsmoke increases during the nighttime, the PM2.5 measurements were conducted between 8pm and midnight using two DustTrak II units and a GPS receiver. While one DustTrak II unit collected mobile PM2.5 concentrations, the other unit collected PM2.5 concentrations at a fixed central site for subsequent calibration and normalization with the background concentrations. This study used a total of 1,290 measurements of PM2.5 collected every 2 minutes (See Fig. 1), of which 596 measurements were collected in June and 694 measurements in July. Therefore, we analyzed the complete set of PM2.5 measurements and also separately by month.

The basic descriptive statistics of the PM2.5 measurements are shown in Table 1. When comparing with the Chilean normative, approximately 12.9%, 11.4%, and 16.4% of the PM2.5 measurements with values in the range of 80 $\mu g/m^3$ to 109 $\mu g/m^3$, 110 $\mu g/m^3$ and 169 $\mu g/m^3$, and greater than 170 $\mu g/m^3$ correspond to environmental alert, pre-emergency, and emergency, respectively [23]. This reveals that over 40% of the PM2.5 measurements surpass the values established by the Chilean normative.

In order to detect outliers in the measurements of PM2.5, a robust measure of dispersion, namely the median absolute deviation (MAD) method, was applied to all PM2.5 measurements, as explained in [24]. Fig. 2 presents the PM2.5 values for the complete mobile sampling campaign. The horizontal red line indicates that PM2.5 concentration values greater than the superior limit of 880.1 $\mu g/m^3$ should be neglected. Thus, a total of 18 measurements of PM2.5, representing 1.4% of the complete sample, were discarded from the PM2.5 measurements.

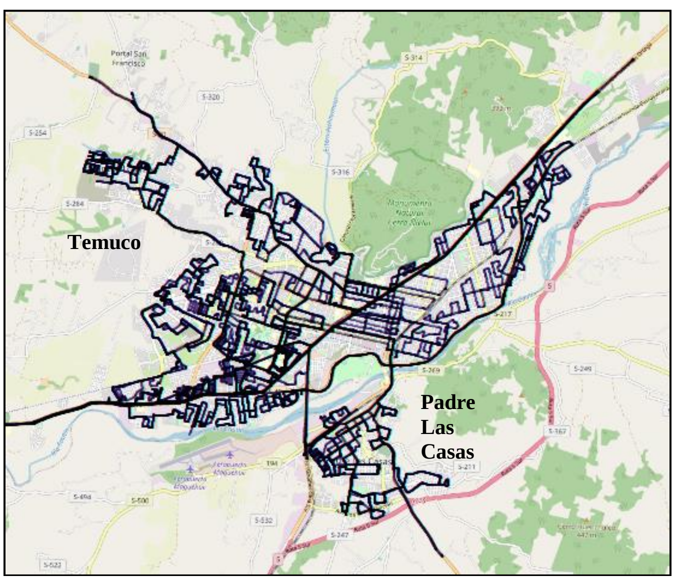


Fig. 1 Mobile sampling routes in Temuco and Padre Las Casas

TABLE I

DESCRIPTIVE STATISTICS OF THE PM2.5 MEASUREMENTS

Month	Number of Measurements	Mean ($\mu g/m^3$)	Standard Deviation ($\mu g/m^3$)	Min ($\mu g/m^3$)	Max ($\mu g/m^3$)
June	596	162.29	219.75	6.25	1483.30
July	694	75.92	68.72	6.71	465.66
Total	1,290	115.82	163.35	6.25	1483.30

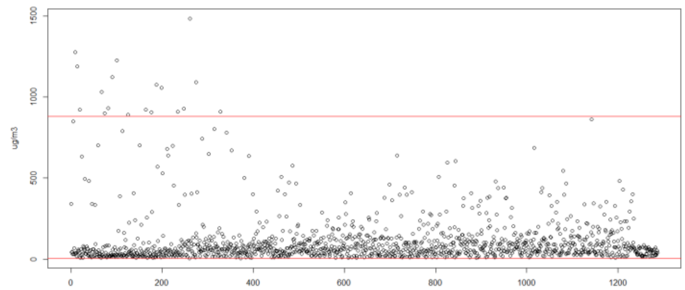
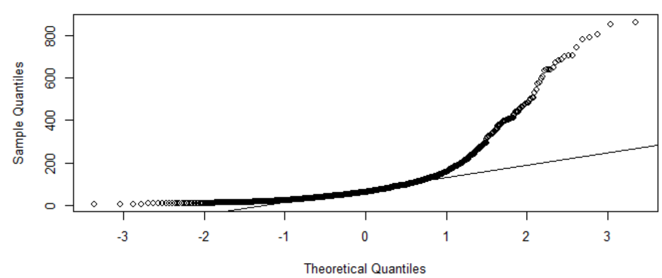
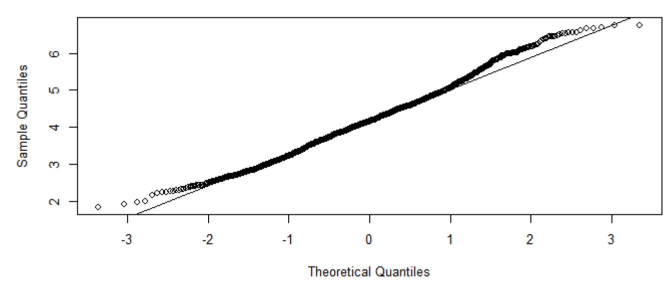


Fig. 2 PM2.5 values for the complete mobile sampling campaign

An exploratory analysis of the complete data set showed that the histogram of the PM2.5 measurements presents a skewness of 2.95 and a kurtosis of 13.45. Therefore, a logarithmic transformation was applied and the coefficient of skewness was reduced to 0.24 and the kurtosis to 2.84. Fig. 3a) and 3b) depict the normal QQ plot with and without the transformation, respectively, suggesting that the data are lognormally distributed. Thus, this log-transformed data is employed in the subsequent analyses.



a) Without transformation



b) With transformation

Fig. 3 Normal QQ plot for all PM2.5 measurements without and with the logarithmic transformation

Additionally, the PM2.5 measurements are analyzed separately per month. A logarithmic transformation was applied to the June and July PM2.5 measurements, yielding a coefficient of skewness equal to 0.08 and 0.17, and kurtosis of 3.31 and 2.87, respectively. The normal QQ plots with the transformation for these measurements collected in June and July are shown Fig. 4 and 5, respectively.

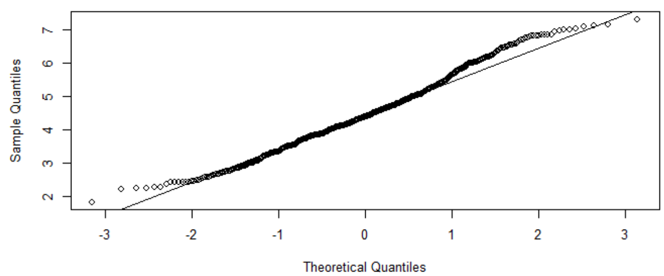


Fig. 4 Normal QQ plot for PM2.5 measurements collected in June with the logarithmic transformation

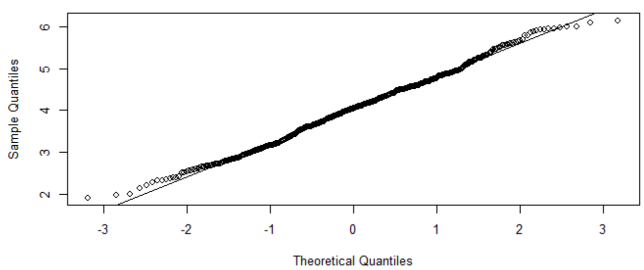


Fig. 5 Normal QQ plot for PM2.5 measurements collected in July with the logarithmic transformation

IV. Methodology

Prior to implementing Kriging, the weights assigned to each estimated value are obtained with a semivariogram model. This model identifies the distance at which the data are no longer autocorrelated (i.e., spatial dependence). Therefore, the semivariogram depends on the distance and also the average sum of squared differences in the values for all pairs of data points that are separated at a certain distance and is computed using (1) [25, 26].

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i + h)\}^2 \qquad (1)$$

Where $\gamma(h)$ is the value of the experimental semivariance at a distance or range h, $Z(x_i)$ are the measured values at location i, $Z(x_i + h)$ are the measured values at location i plus a distance h, and $N(h)$ is the total number of pairs separated by a distance h.

The main values of the semivariogram are the nugget, sill, and range, as shown in Fig. 6. The nugget is the value at which the curve intercepts the Y axis, the sill is the maximum value of the curve representing the theorical sample variance, and the range h is the value along the X axis at which a certain threshold is reached (i.e., a measured point may not contribute to the estimation value at an unmeasured location if located beyond the distance given by the range) [27].

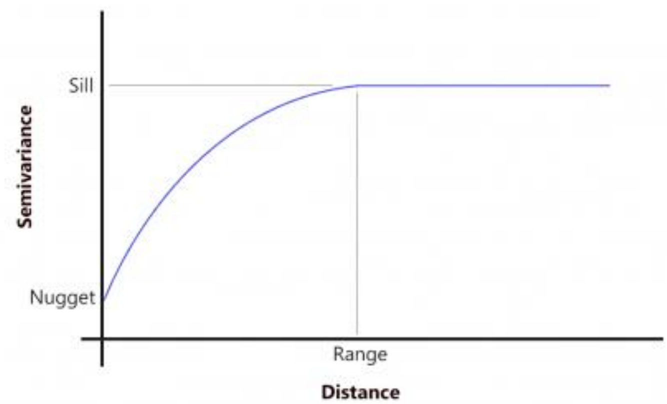


Fig. 6 Example of a semivariogram and its components (nugget, sill, and range) Source: Reference [28]

Kriging is a geostatistical technique that interpolates values at unmeasured locations based on measured locations. This technique has the advantage of providing an error estimate for unmeasured locations, and uses spatial correlations (i.e., semivariogram) to explain variations in the predicted surface [24]. The weights associated to each value are computed based on the semivariogram model. The ordinary Kriging will be employed in this study, as in [26, 29], which assumes that the constant mean is unknown. The ordinary Kriging is based on the general expression given by (2).

$$Z^*(x_o) = \sum_{i=1}^{n} \lambda_i Z(x_i) \qquad (2)$$

Where $Z^*(x_o)$ is the estimated value at predicted location x_o, λ_i is the weighting coefficient at location i, $Z(x_i)$ are the measured values at location i, and n is the number of measured values.

V. Results

A. Semivariogram

Prior to implementing the geostatistical tool, Kriging, a semivariogram was generated to obtain the spatial dependence of the data points through the fitting of a model using R software. A direction tolerance of 45 degrees with a spherical model presented the best spatial continuity for the complete set of PM2.5 measurements, as shown in Fig. 7. Fig. 8 and 9 present the semivariograms for the PM2.5 measurements collected in June and July, respectively. The PM2.5 measurements collected in June and July presented the best fit with Gaussian and circular models and direction tolerances of 45 and 135 degrees, respectively.

Table II presents the values of the nugget, sill, and range of these semivariogram for the complete set of PM2.5 measurements and also for the June and July measurements, separately. These values are employed in the subsequent analysis with Kriging.

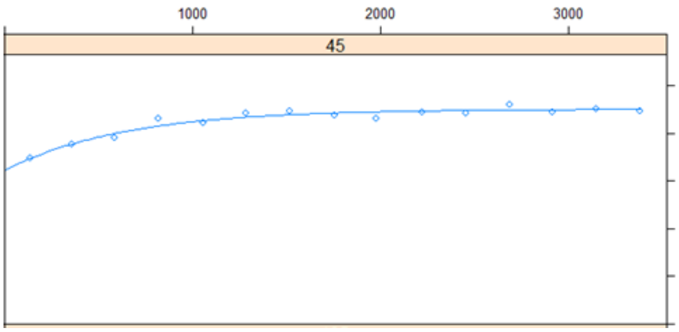


Fig. 7 Semivariogram with a direction tolerance of 45 degrees for the Spherical model using all PM2.5 measurements

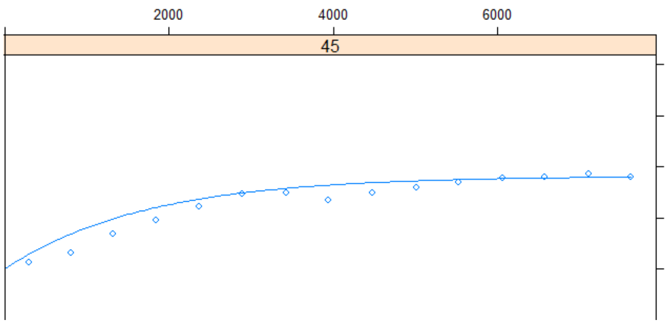


Fig. 8 Semivariogram with a direction tolerance of 45 degrees for the Gaussian model using PM2.5 measurements collected in June

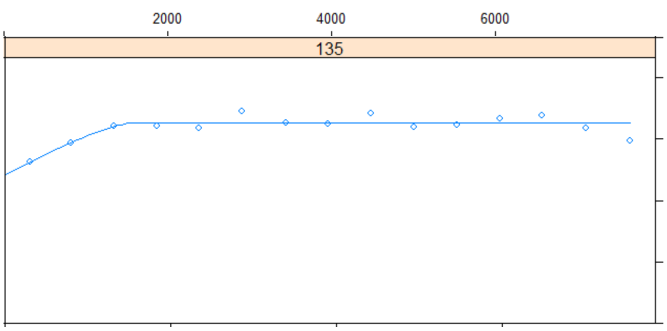


Fig. 9 Semivariogram with a direction tolerance of 135 degrees for the Circular model using PM2.5 measurements collected in July

TABLE II
SEMIVARIOGRAM COMPONENT VALUES FOR THE COMPLETE SET, AND FOR JUNE AND JULY PM2.5 MEASUREMENTS

Semivariogram Components	PM2.5 Measurements		
	Whole period	June	July
Nugget (m)	0.6	0.5	0.65
Sill (m)	0.8	1.3	0.5
Range (m)	1,200	6,000	2,000

B. *Ordinary Kriging*

Subsequently, the Ordinary Kriging employed the components obtained from the semivariogram model. Fig. 10 shows the estimated surface with the spatial variability using all PM2.5 measurements, where the lighter (darker) colors indicate high (low) concentrations of PM2.5. The inverse log-transformation of these estimates reveal that PM2.5 concentrations could reach values near 150 μg/m^3 in some areas. Fig. 11 and 12 present the spatial variability of PM2.5 concentrations in Temuco and PLC for measurements collected in June and July, respectively. These figures show that PM2.5 concentrations in June varied between 20 and 665 μg/m^3, whereas, in July, PM2.5 concentrations ranged from 30 to 140 μg/m^3.

Kriging also has another output that shows the reliability of the PM2.5 estimates, yielding a Kriging variance for each measured point. This reliability depends on the selected spatial model (e.g., Spherical, Gaussian, or Circular) in the semivariagram. Fig. 13, 14, and 15 show the Kriging variance with the reliability level of the estimates for all PM2.5 measurements, and those PM2.5 concentrations measured in June, and July, respectively. In these figures, lighter colors are more reliable than darker colors. Note that lighter colors tend to coincide with the PM2.5 measurement routes, while no PM2.5 measurements were collected in areas with the color magenta, showing less reliable PM2.5 estimates.

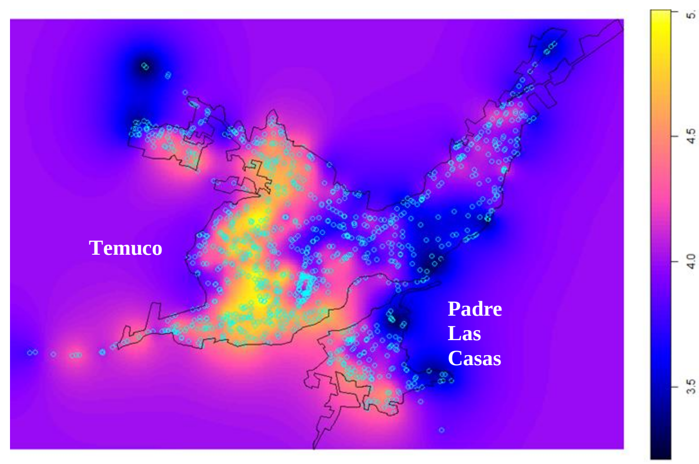


Fig. 10 Estimated surface of all PM2.5 measurements in Temuco and PLC

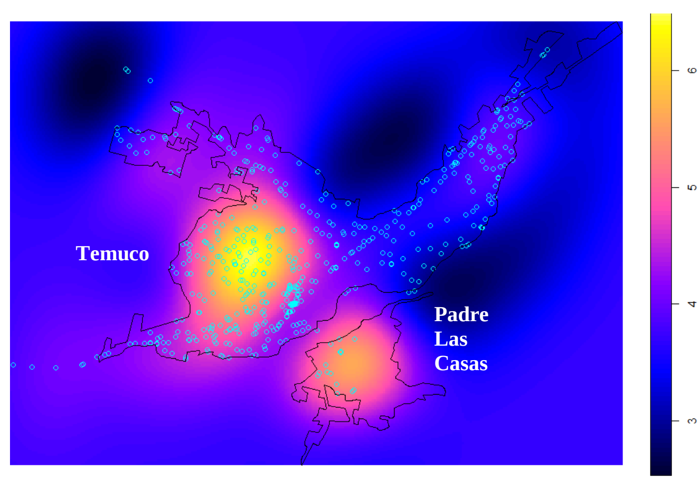


Fig. 11 Estimated surface of PM2.5 measurements collected in June in Temuco and PLC

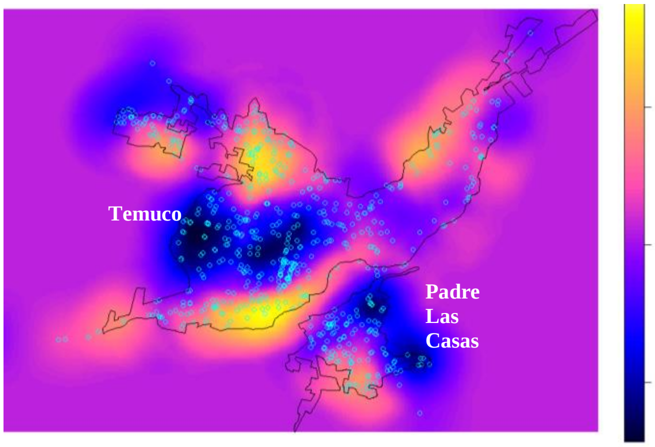


Fig. 12 Estimated surface of PM2.5 measurements collected in July in Temuco and PLC

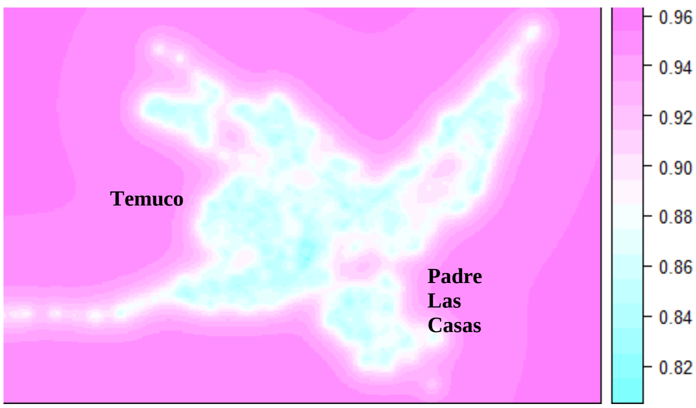


Fig. 13 Kriging variance of all estimated PM2.5 measurements in Temuco and PLC

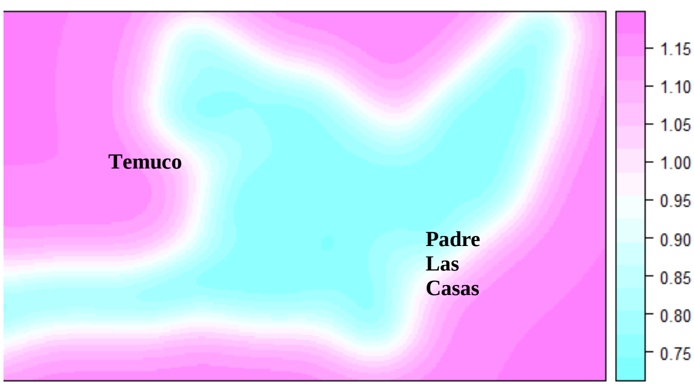


Fig. 14 Kriging variance of estimated PM2.5 measurements in June in Temuco and PLC

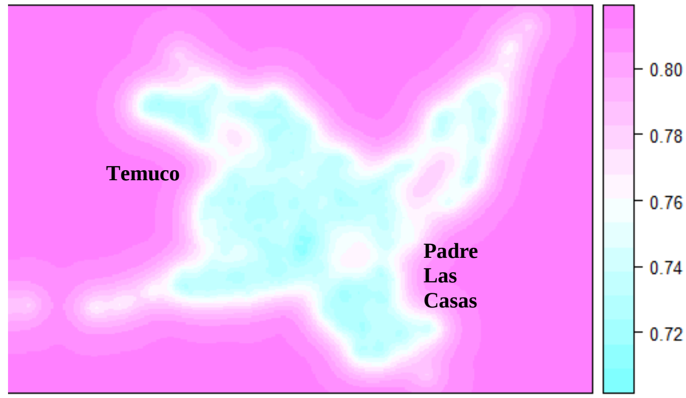


Fig. 15 Kriging variance of estimated PM2.5 measurements in July in Temuco and PLC

VI. Conclusions

In this study, we employed PM2.5 concentrations that were collected in a mobile campaign in the conurbation of Temuco and Padre Las Casas (PLC) in Chile during the winter of 2016. The results of this study show the prediction of PM2.5 concentrations at unmeasured locations in this conurbation using the geostatistical method, namely Ordinary Kriging. Overall, the results of this study suggest that higher PM2.5 concentrations were collected in Temuco than in PLC, similar to the results of [30]. Therefore, local authorities should implement environmental measures to reduce PM2.5 concentrations in these areas, and thus, improve the air quality and the health of the community.

When considering the complete set of PM2.5 measurements, we observed high PM2.5 concentrations throughout the western part of Temuco and only part of PLC. However, differences exist when examining the PM2.5 measurements by month (June and July). While in June high PM2.5 values (as high as 665 $\mu g/m^3$) are more concentrated toward the west of Temuco (Amanecer, Estadio Municipal, Av. Alemania, Barrio Inglés) and large part of PLC, high PM2.5 concentrations are clustered in the north, north-east, and south of Temuco and south-west of PLC in July, as in [20]. Further analysis of the collections routes for each month are required to analyze the difference in the PM2.5 concentrations. Perhaps there were less routes conducted in July in the center of Temuco, and this generated higher concentration values of PM2.5 toward other parts of the city. Although higher PM2.5 concentrations were predicted in the June measurements, less reliable PM2.5 estimates were obtained in this month using the Kriging variance. This result requires further investigation.

Future research should include the prediction of PM2.5 concentrations by hour and day of the week to determine the spatial variability at smaller temporal units. Additionally, meteorological and environmental conditions should be studied to comprehend their impact on the estimated PM2.5 concentrations in space and time. Finally, additional mobile campaigns are need in Temuco and PLC to update the PM2.5 measurements employed in this study.

ACKNOWLEDGMENTS

The authors would like to thank to Dr. Pablo Ruiz from Public Health School at Universidad de Chile and Dr. María Elisa Quinteros from Universidad de Talca for providing the mobile PM2.5 measurements.

REFERENCES

[1] M. Brauer, G. Freedman, J. Frostad, A. van Donkelaar, R. V. Martin, F. Dentener et al., "Ambient air pollution exposure estimation for the global burden of disease 2013," *Environmental Science & Technology*, vol. 50, no. 1, pp. 79–88, 2016.

[2] J. Apte, M. Brauer, A. Cohen, M. Ezzati, and C. Pope, "Ambient PM2.5 reduces global and regional life expectancy", *Environmental Science & Technology Letters*, vol. 5, no. 9, pp. 546-551, 2018.

[3] J. Anderson, J. Thundiyil, and A. Stolbach, "Clearing the air: A review of the effects of particulate matter air pollution on human health", *Journal of Medical Toxicology*, vol. 8, pp. 166-175, 2012.

[4] A. M. Villalobos, F. Barraza, H. Jorquera, and J. J. Schauer, "Wood burning pollution in southern Chile: PM2.5 source apportionment using CMB and molecular markers," *Environmental Pollution*, vol. 225, pp. 514–523, 2017.

[5] Instituto Nacional de Estadística, INE, "Censo 2017." [Online]. Available: https://www.censo2017.cl/

[6] M. E. Quinteros, S. Lu, C. Blazquez, J. P. Cardenas-R, X. Ossa, J.-M. Delgado-Saborit et al., "Use of data imputation tools to reconstruct incomplete air quality datasets: A case-study in Temuco, Chile," *Atmospheric Environment*, vol. 200, pp. 40 – 49, 2019.

[7] F. Muñoz, and M. Carvalho, "Effect of exposure time to PM10 on emergency admissions for acute bronchitis," *Cadernos de Saúde Pública*, vol. 25, no. 3, 2009.

[8] S. Barrios Casas, F. Peña-Cortés, and S. Osses Bustingorry, "Effects for particles material atmospheric pollution on acute respiratory diseases in under 5 years of age," *Ciencia y Enfermería*, vol. X, no. 2, pp. 21-29, 2004.

[9] S. Castellanos González, "Validation of PM2.5 in Bogota with a predictive method based on Kriging", 2019. [Online]. Available: https://www.academia.edu/download/65071012/Validation_of_PM2.5_in_Bogota_with_a_predictive_method_based_on_Kriging.pdf

[10] H. Rodriguez, "Application of interpolation methods and geostatistical modeling in the evaluation of air quality in Bogotá D.C.", Universidad Militar Nueva Granada, 2014. [Online]. Available: http://repository.unimilitar.edu.co/handle/10654/13446

[11] L. Knibbs, A. van Donkelaar, R. Martin, M. Bechle, M. Brauer, D. Cohen, C. Cowie, M. Dirgawati, Y. Guo, I. Hanigan, F. Johnston, G. Marks, J. Marshall, G. Pereira, B. Jalaludin, J. Heyworth, G. Morgan, and A. Barnett, "Satellite-Based Land-Use Regression for Continental-Scale Long-Term Ambient PM2.5 Exposure Assessment in Australia," *Environmental Science & Technology*, vol. 52, no. 21, pp. 12445-12455, 2018.

[12] T. Shi, Y. Hu, M. Liu, C. Li, C. Zhang, and C. Liu, "Land use regression modelling of PM2.5 spatial variations in different seasons in urban areas," *Science of the Total Environment*, vol. 743, no. 15, 140744, 2020.

[13] X. Hu, J. Belle, X. Meng, A. Wildani, L. Waller, M. Strickland, and Y. Liu, "Estimating PM2.5 concentrations in the conterminous United States using the random forest approach," *Environmental Science & Technology*, vol. 51, no. 12, pp. 6936-6944, 2017.

[14] G. Chen, S. Li, N. Hamm, W. Cao, T. Li, J. Guo, H. Ren, M. Abramson, and Y. Guo, "A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information," *Science of the Total Environment*, vol. 636, pp. 52-60, 2018.

[15] L. Rodríguez-Camargo, R. Sierra-Parada1, and L. Blanco-Becerra "Análisis espacial de las concentraciones de PM2,5 en Bogotá según los valores de las guías de la calidad del aire de la Organización Mundial de la Salud para enfermedades cardiopulmonares, 2014-2015," *Biomédica*, vol. 40, pp. 137-152, 2020.

[16] K. Shukla, P. Kumar, G. Mann, and M. Khare, "Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi," *Sustainable Cities and Society*, vol. 54, 101997, 2020.

[17] C. Magri, "Spatial and temporal patterns of PM2.5 In Santiago Chile," ESRI User Conference, 2017 [Online]. Available: http://proceedings.esri.com/library/userconf/proc17/papers/595_336.pdf

[18] M. Xu, H. Sbihi, X. Pan, and M. Brauer, Modifiers of the effect of short-term variation in PM2.5 on mortality in Beijing, China," *Environmental Research*, vol. 183, 109066, 2020.

[19] L. Zhang, J. An, M. Liu, Z. Li, Y. Liu, L. Tao, X. Liu, F. Zhang, D. Zheng, Q. Gao, X. Guo, and Y. Luo, "Spatiotemporal variations and influencing factors of PM2.5 concentrations in Beijing, China," *Environmental Pollution*, vol. 262, 114276, 2020.

[20] C. Johansson, G. Olivares, and L. Gidhagen, "Characterisation and source apportionment of particulate matter in two urban areas of Chile," ITM-report 169, Institutionen för tillämpad miljövetenskap, 2007, [Online]. Available: https://www.slb.nu/slb/rapporter/pdf8/itm2007_169.pdf

[21] A. Villalobos, F. Barraza, H. Jorquera, and J. Schauer, "Wood burning pollution in southern Chile: PM2.5 source apportionment using CMB and molecular markers," Environmental Pollution, vol. 225, pp. 514-523, 2017.

[22] H. Jorquera, F. Barraza, F., J. Heyer, G. Valdivia, G., L. Schiappacasse, and L. Montoya, "Indoor PM2.5 in an urban zone with heavy wood smoke pollution: the case of Temuco, Chile," Environmental Pollution, vol. 236, 477e487, 2018.

[23] M. del Medio Ambiente, "Decreto nº12 que establece norma primaria de calidad ambiental para material particulado fina respirable," 2011. [Online]. Available: https://www.leychile.cl/Navegar?idNorma=1025202

[24] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764-766, 2013.

[25] D. O'Sullivan, and D. J. Unwin, *Geographic Information Analysis*, New Jersey, John Wiley & Sons, Inc., 2003.

[26] F. Moral, P. Alvarez, and J. Canito, "Mapping and hazard assessment of atmospheric pollution in a medium sized urban area using the Rasch model and geostatistics techniques," *Atmospheric Environment*, vol. 40, no. 8, pp. 1408-1418, 2006.

[27] P. Burrough, and R. McDonnell, *Principles of Geographical Information Systems*, New York, Oxford University Press, 1998.

[28] GIS Geography, "Semi-Variogram: Nugget, Range and Sill", [Online]. Available: https://gisgeography.com/semi-variogram-nugget-range-sill/

[29] Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 1997.

[30] C. Blazquez, and E. Montero, "Spatial and aspatial clustering analysis of PM2.5 concentrations in Temuco, Chile using mobile measurements," 18th LACCEI International Multi-Conference for Engineering, Education Caribbean Conference for Engineering and Technology, Latin American and Caribbean Consortium of Engineering Institutions, 2020.