SPATIAL AND TEMPORAL PATTERNS OF $PM_{2.5}$ IN SANTIAGO CHILE

INTRODUCTION

The health impacts of inhaling particulate matter (PM) are well documented. Short term (hours, days) and long term (months, years) exposure leads to cardiovascular and respiratory diseases as well as lung cancer. Morbidity increases in patients with preexisting lung or heart diseases, the elderly and children. In children long term exposure to PM affects lung development and lung function (WHO 2006, WHO 2013, Green & Sanchez, 2013).

Particulate matter is a common air pollutant that consists of a mix of solid and liquid particles that varies by location. It is usually described by its mass concentration (micrograms per cubic meter, $\mu g/m^3$) and classified according its particle diameter: PM_{10} refers to particles with diameters less than 10 microns (μm), and $PM_{2.5}$ to particles with diameters less than 2.5 μm which include ultrafine particles that have diameters smaller than 0.1 μm. European studies suggest that $PM_{2.5}$ constitutes 50 to 70% of PM_{10} (WHO Europe, 2013). Particles with diameters between 0.1 and 1 μm can remain in the atmosphere for days or weeks and can be carried long distances through the atmosphere (WHO Europe, 2006).

Particulate matter is a mix of physical and chemical particles that can be emitted directly from a source (primary particles) or formed in the atmosphere through chemical reactions of gaseous pollutants (such as sulfur dioxide, hydroxyl radical and nitrogen oxides) in the presence of high solar radiation (secondary particles).

Primary sources of PM are both man-made and natural. Man-made sources are combustion engines, solid fuel combustion for energy production for households and industry, and other activities such as construction, mining, agriculture, residential wood or coal burning for heating or cooking, pavement erosion due to traffic and the wear-down of brakes and tires. Natural sources of PM are soil, dust resuspension, forest and grassland fires to name a few.

The objective of this project is to examine the spatial and temporal patterns of hourly $PM_{2.5}$ concentration in Santiago, Chile using 3D kriging and the ArcGIS Space Time Pattern Mining toolbox. The goal is to identify areas of concentration and levels of exposure that exceed recommended levels. The project will also explore seasonal variations of the concentration throughout the city and the relationship between $PM_{2.5}$ and environmental factors (such as wind speed and direction, relative humidity, air temperature and elevation).

LITERATURE REVIEW

In 2015, the British Medical Journal published two papers (Shah et al. 2015 and Power et al. 2015) that showed effects of bad air quality on the brain. Shah et al. analyzed 94 studies (6.2 million cases of stroke in 28 countries) and found a significant positive correlation between air pollution levels during the seven days before hospitalizations or deaths due to stroke, with the level of exposure of on the day of the stroke being particularly significant. Power et al. conducted a study of 71,271 women between 57 to 85

years of age and found that symptoms of high anxiety were significantly more common among women exposed to high concentrations of $PM_{2.5}$.

These two studies provide evidence that exposure to $PM_{2.5}$ promotes inflammatory processes in the body and increases oxidative stress by particles that have been deposited deep in the lungs. These particles are then carried by the blood to the heart, brain and other organs. Sherman (2015) from the Dana Foundation mentions that other studies have shown that cognitive function is poorest among adults that live in areas with higher $PM_{2.5}$ concentrations and that the effect is equivalent to almost 2 years of brain aging.

The WHO Report on Health Effects of Particulate Matter (WHO, 2013) states that $PM_{2.5}$ is a stronger risk factor for mortality that PM_{10}, especially long term exposure. Long-term exposure to $PM_{2.5}$ is associated with an increase in the long-term risk of cardiopulmonary mortality of 6–13% per 10 $\mu g/m^3$ of $PM_{2.5}$. The black carbon portion of $PM_{2.5}$, which is the result of incomplete combustion, is considered a known carcinogen and a major contributor to global climate change because it absorbs light and generates heat in the atmosphere (Green, 2013, Riojas-Rodríguez et al. 2016). The WHO (2013) report also estimates that approximately 3% of cardiopulmonary and 5% of lung cancer deaths are attributable to PM globally. The same report defines the guidelines for acceptable levels of $PM_{2.5}$ as a maximum of 10 $\mu g/m^3$ for the annual average and 25 $\mu g/m^3$ for the 24-hour mean (not to be exceeded for more than 3 days/year).

Studies have shown that $PM_{2.5}$ concentration has a distinct pattern during the day, showing a bimodal peak of concentration between 7 and 8 AM and another peak between 7 to 11 PM, with a minimum concentration around noon (Zirui et al. 2015, Gramsch et al. 2014, Perez & Gramsch 2016). Meteorological variables such as wind speed and air temperature have negative correlations that are stronger than the one for relative humidity on $PM_{2.5}$ concentrations when these variables are compared on a daily basis (i.e. daily averages of $PM_{2.5}$ and daily meteorological values). Wind speed has the most effect and shows an inverse correlation, especially when the topography of the area is considered as well (Zirui et al., 2015, Jhun et al., 2013, Li et al., 2015).

STUDY BACKGROUND

Santiago, shown in Figure 1 (33.5 S, 70.5 W, 520m/1,706ft above mean sea level), is the capital of Chile and has a population of 7.3 million people. The city is surrounded by two mountains ranges: the Andes to the east and the Coastal range to the west, both with peaks higher than 2,000 meters above sea level, which makes air circulation difficult, especially in winter (when a low altitude thermal inversion layer forms and keeps pollutants close to the surface). The city also has 4.7 million vehicles and 70% of the country´s total mineral and agricultural industries are concentrated in the surrounding area (MMA 2014 and MMA 2015).

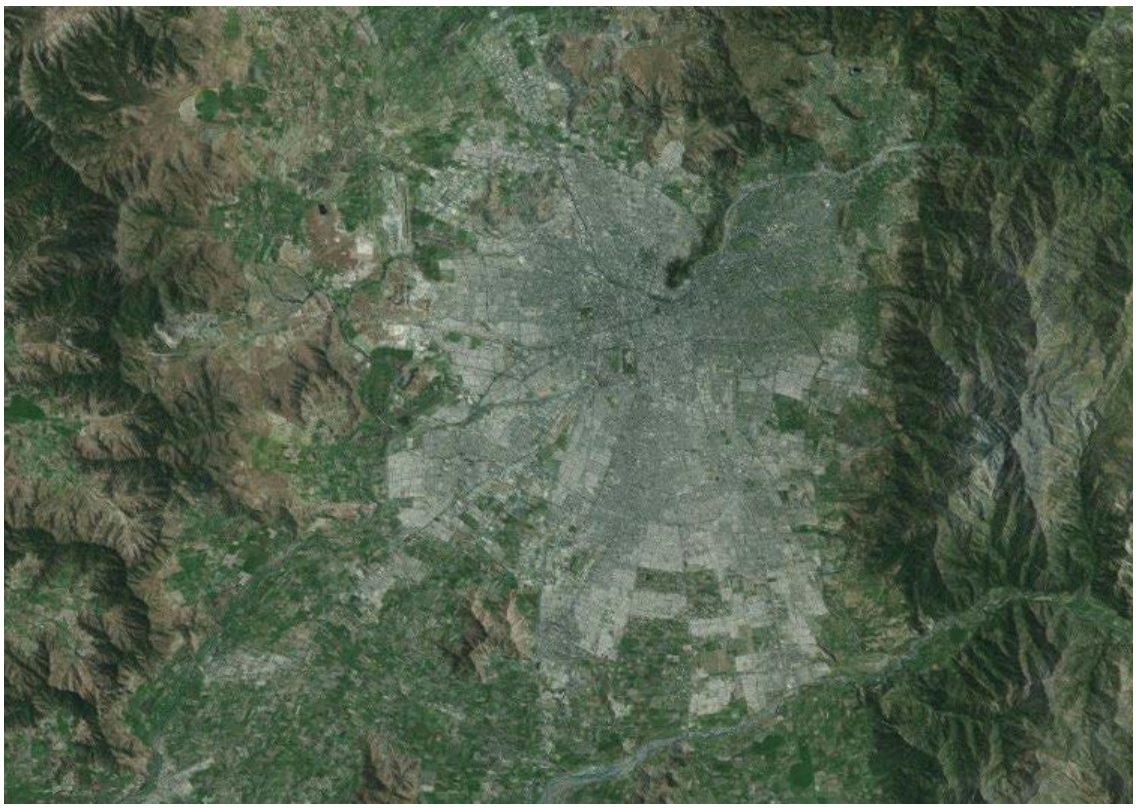


Figure 1: Santiago and surrounding areas.

Since 1997 Chile has implemented decontamination plans that have reduced the levels of air pollution in Santiago. $PM_{2.5}$ levels declined for the first 13 years and have been stable since 2009. However, Santiago persistently exceeds the daily and annual limits for $PM_{2.5}$ concentrations, which are 50 and 20 $\mu g/m^3$, respectively, as defined in Chilean law (D.S. N°12/2011). The WHO guidelines mentioned above are also persistently exceeded (which are 25 $\mu g/m^3$ for the daily mean and 10 $\mu g/m^3$ for the annual mean). In 2005, the Chilean Ministry of the Environment and the World Health Organization estimated that, at a national level, 4,000 premature deaths are the consequence of atmospheric pollution, costing the country 670 million dollars in medical expenses and loss of productivity. These estimates were based on a paper by Cifuentes et al. (2005).

A press release by Gramsch (2014) and a study by the Chilean Ministry of the Environment (MMA, 2014) indicate that the major sources of $PM_{2.5}$ in Santiago are residential burning of wood (22%, but reaching 30% in winter time), public transportation (8%), other vehicles (36%, including 3% attributable to private vehicles) and industry and agriculture (33%). Gramsch et al. (2014) also reports that between 10% and 20% of $PM_{2.5}$ in Santiago is Black Carbon.

The current policy in Santiago implements emergency measures based on PM_{10} levels, but this was expected to change to $PM_{2.5}$ in 2017 when a new decontamination plan (called “Santiago Respira” or “Santiago Breathes”) would become law. An average of 24 hourly measurements is used to determine if pollution levels are high enough to implement mitigation measures. The new plan has a threshold of 79

$\mu g/m^3$ for $PM_{2.5}$ before emergency measures are implemented, whereas the current plan is based on PM_{10}.

EXPLORATORY ANALYSIS

Hourly data is available from SINCA (the National Information System on Air Quality) for $PM_{2.5}$ concentrations and weather conditions for 11 measurement stations distributed throughout Santiago, as shown in Figure 2.

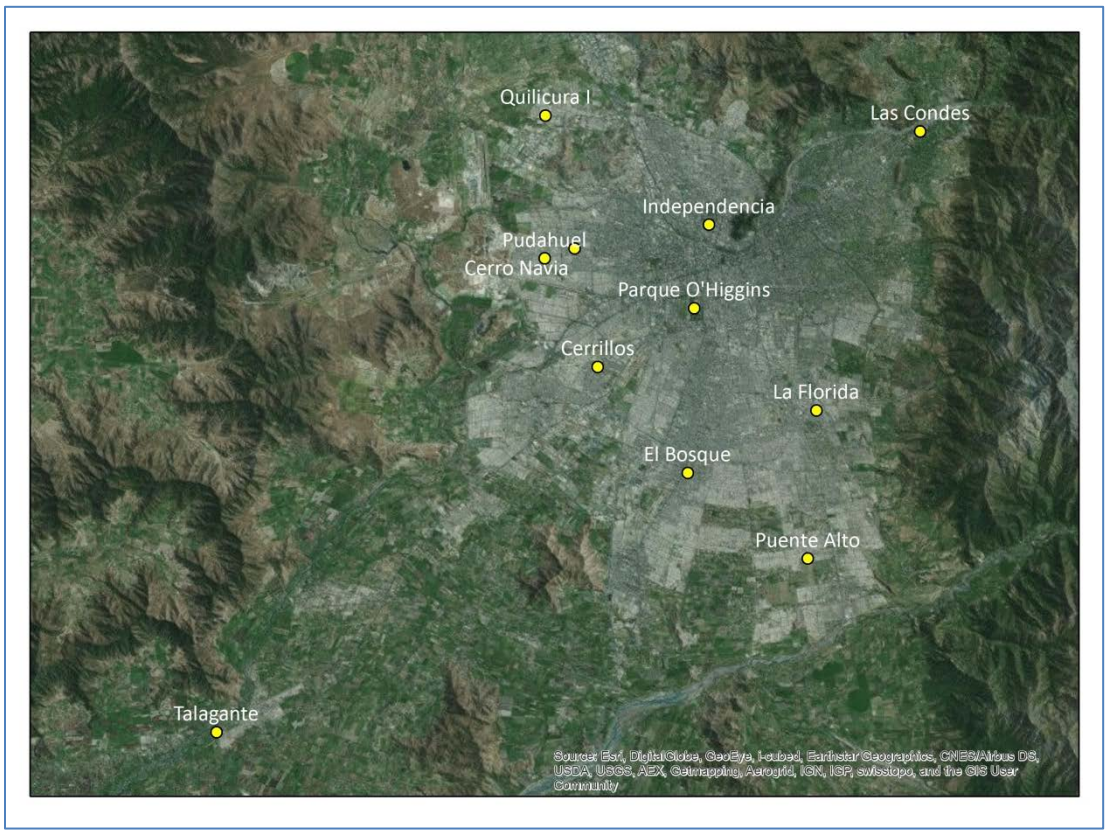


Figure 2: Location of the measurement station in Santiago, Chile

Table 1 shows the characteristics of the measurement stations:

Station Name	Station Code	Projection	Easting	Northing	Elevation (meters)
Independencia	D11	UTM 19	346483.7	6300685.5	565
La Florida	D12	UTM 19	352507.8	6290309.8	601
Las Condes	D13	UTM 19	358305	6305906	789
Parque O´Higgins	D14	UTM 19	345662.1	6296020.7	535
Pudahuel	D15	UTM 19	337310	6298804.8	489
Cerrillos	D16	UTM 19	340266.3	6292741.9	509
El Bosque	D17	UTM 19	345307.4	6286812.7	578
Cerro Navia	D18	UTM 19	338974.8	6299356.2	496
Puente Alto	D27	UTM 19	352015.9	6282013.3	660
Talagante	D28	UTM 19	318942.6	6272305.8	309
Quilicura	D29	UTM 19	337348.8	6306794.1	482

Table 1: Measurement Stations

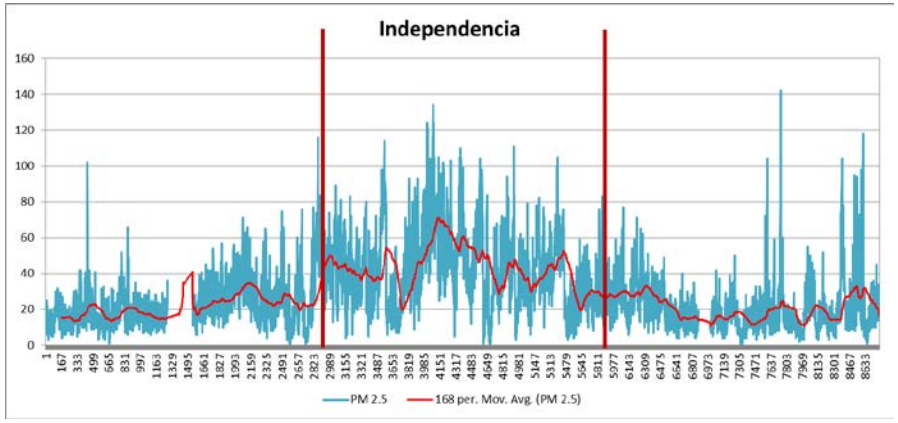
The data analyzed in this project are hourly measurements starting at 00:00AM on January 1, 2016 and ending at 11:00PM on December 31, 2016. For each station there are 8,784 observations (2016 was a leap year), giving a total of 96,623 points of which 3,892 (4%) correspond to missing values. $PM_{2.5}$ summary statistics of the hourly measurements recorded at station for the period of study are shown in Table 2.

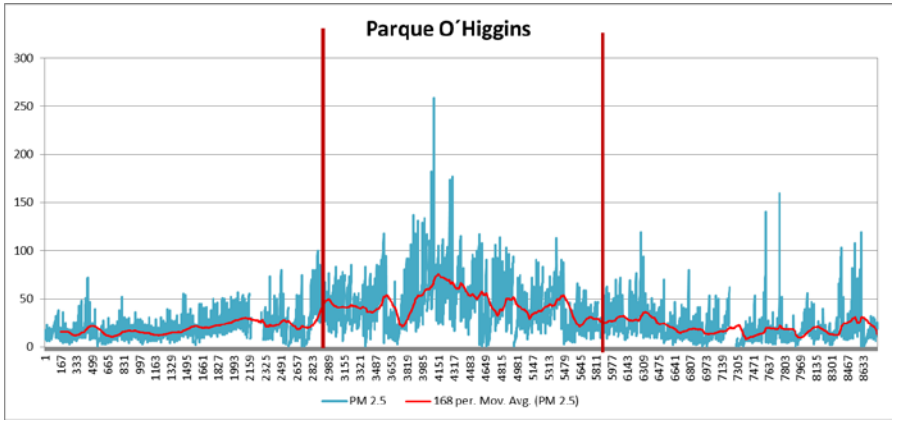
	D11	D12	D13	D14	D15	D16	D17	D18	D27	D28	D29
Min	1	0	0	0	1	0	0	0	0	0	0
Max	142	234	121	259	580	447	562	538	306	219	215
Count	8305	8411	8074	8520	8292	8596	8375	8592	8649	8384	8490
Variance	387.49	480.66	289.58	481.06	1063.37	818.26	1017.52	1258.81	406.29	610.34	431.25
Std. Dev.	19.68	21.92	17.02	21.93	32.61	28.61	31.90	35.48	20.16	24.71	20.77
Average Jan-Dec	29.02	29.20	25.98	28.66	32.04	31.08	34.72	32.91	25.94	24.49	28.05
Average Jan-April	22.33	22.79	23.70	20.22	21.74	21.09	24.14	19.90	22.64	13.97	21.80
Average May-Aug	42.67	43.70	32.70	44.45	51.58	51.87	55.79	56.15	38.22	43.99	41.69
Average: Sept-Dec	20.87	20.02	20.73	20.52	20.93	19.42	22.84	21.71	17.11	14.02	19.93

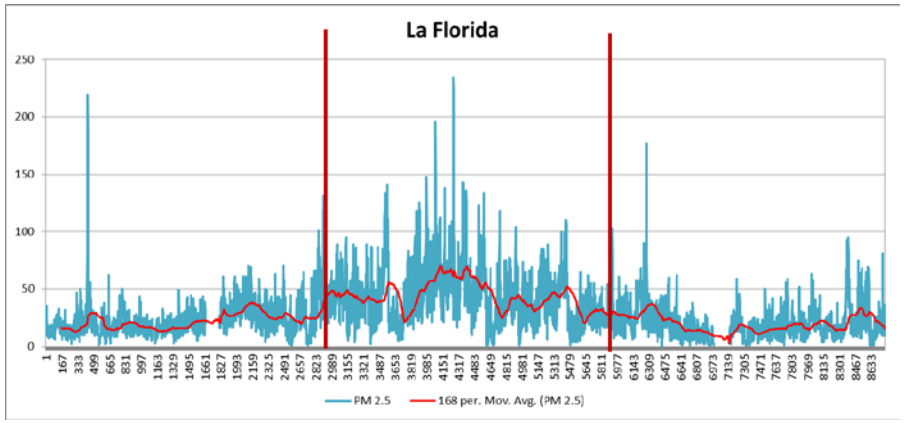
Table 2: Summary Statistics for hourly PM 2.5 concentration per station for study period.

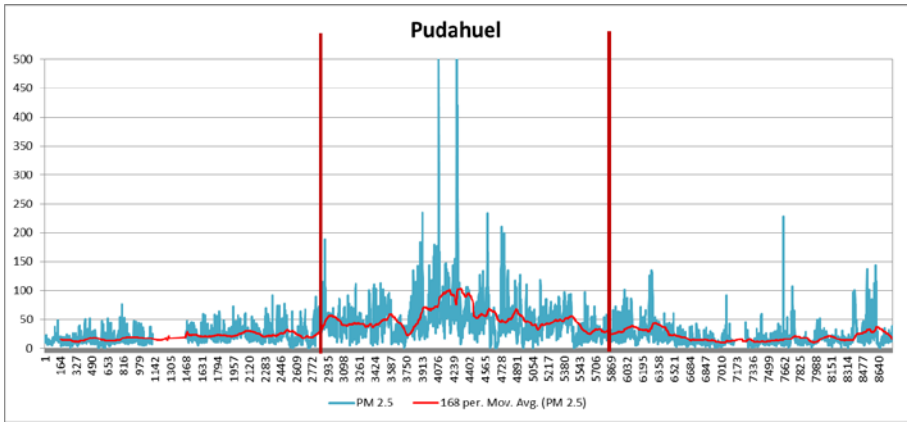
A time series plot for each station is shown in the graphs below (Figure 3). A moving average of 168 hours (one week) is shown in red and indicates variation in $PM_{2.5}$ concentration between warmer and cooler months. The year was divided into three periods of 4 months each: January 1 to April 30, May 1 to August 31 (which correspond to the months when emergency measurements are implemented) and September 1 to December 31. These limits are depicted as vertical red lines on each graph. Summary statistics for each of the three periods are also shown in Table 2.

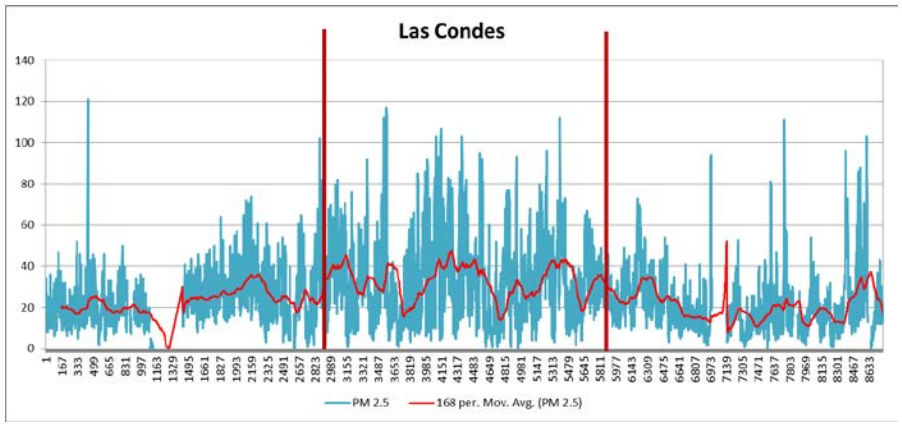
Figure 3: Time series graphs for each station

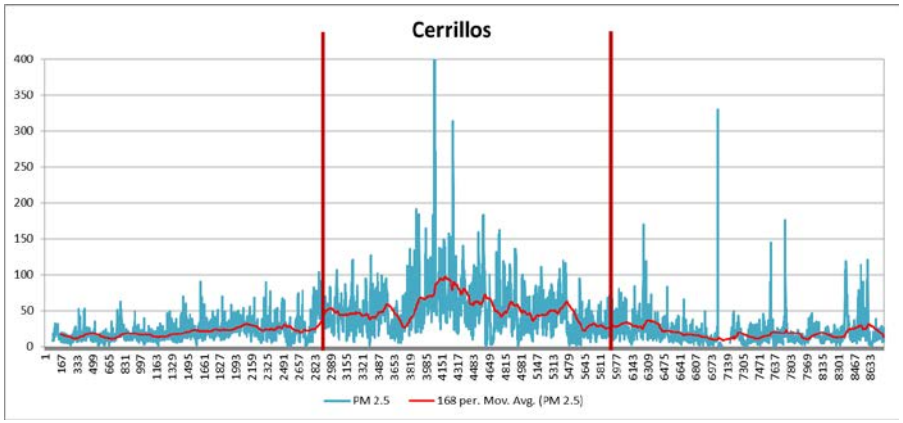


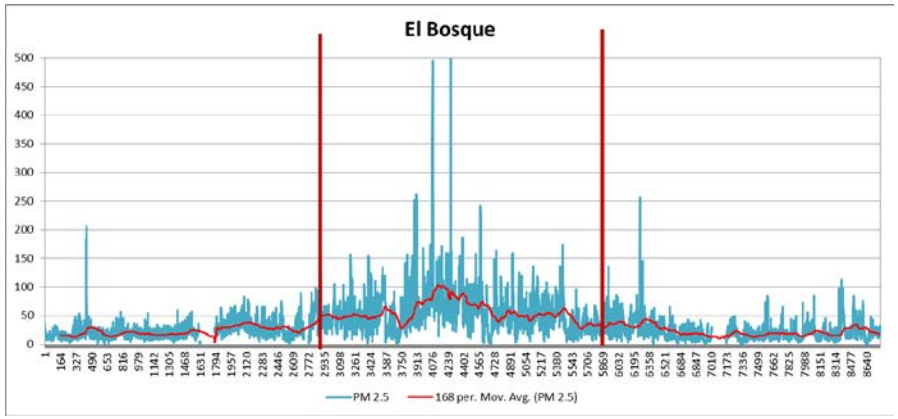


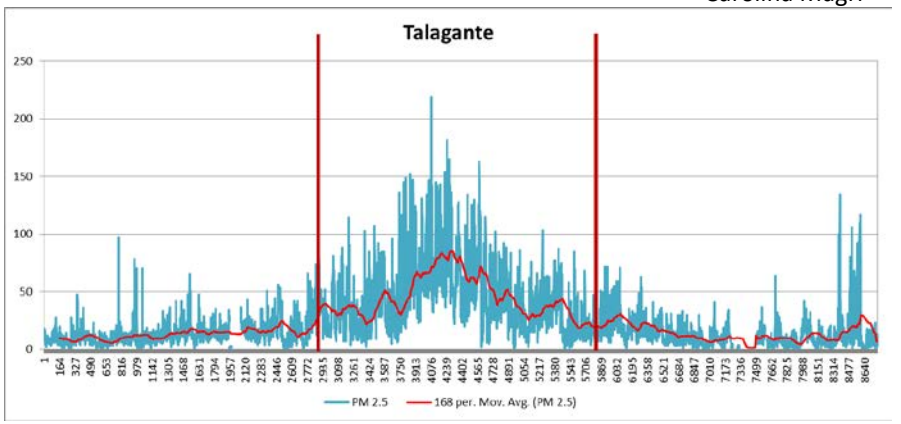


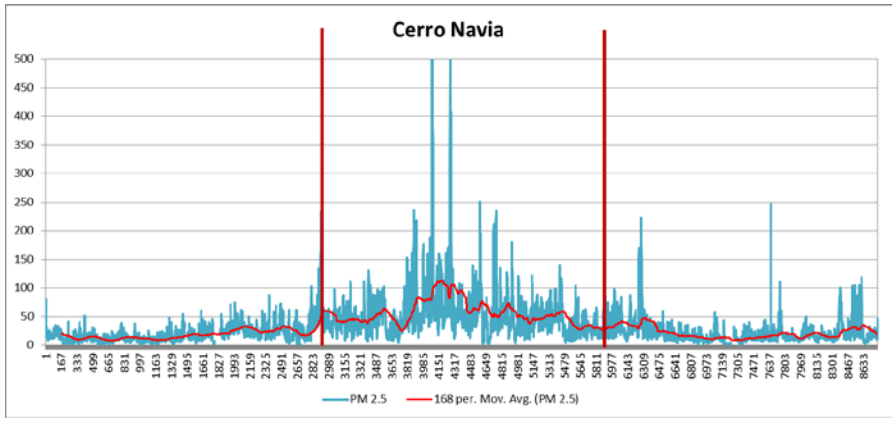


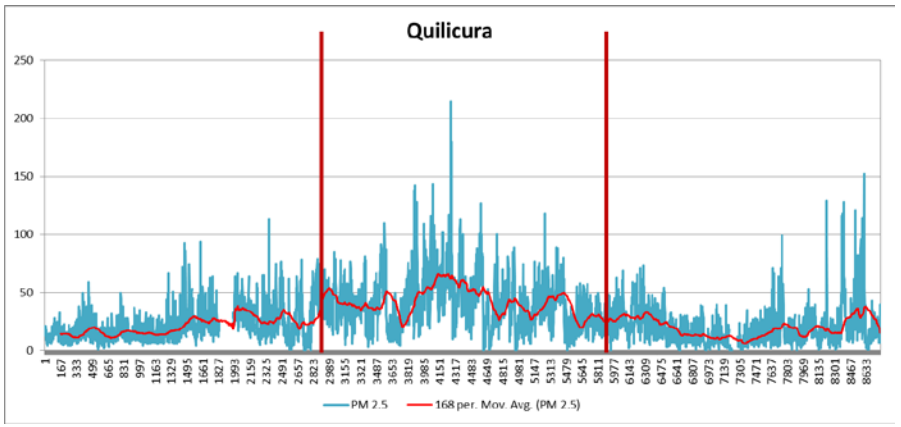


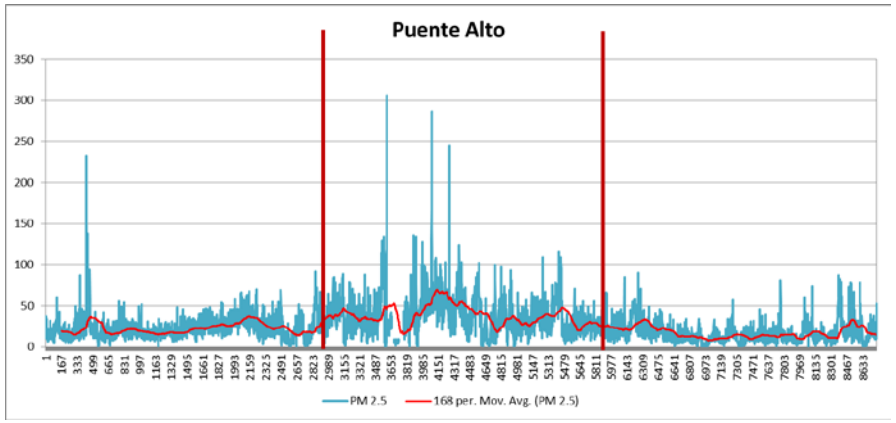



El Bosque
PM 2.5
168 per. Mov. Avg. (PM 2.5)


Talagante
PM 2.5
168 per. Mov. Avg. (PM 2.5)


Cerro Navia
PM 2.5
168 per. Mov. Avg. (PM 2.5)


Quilicura
PM 2.5
168 per. Mov. Avg. (PM 2.5)

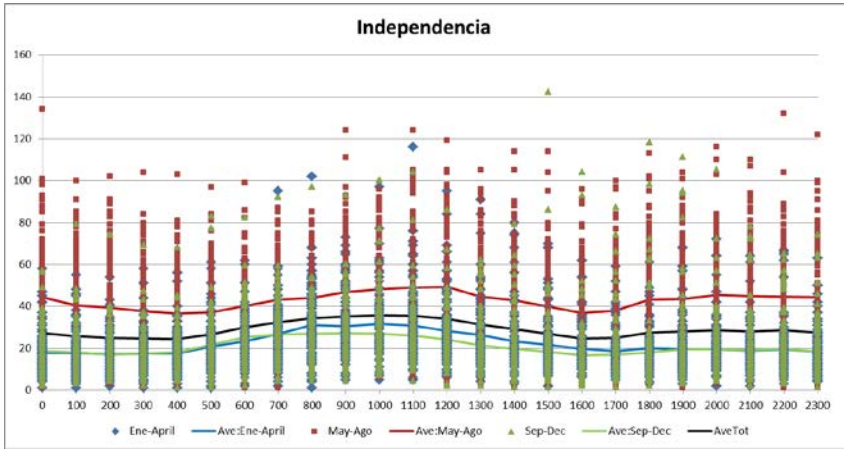

Puente Alto
PM 2.5
168 per. Mov. Avg. (PM 2.5)

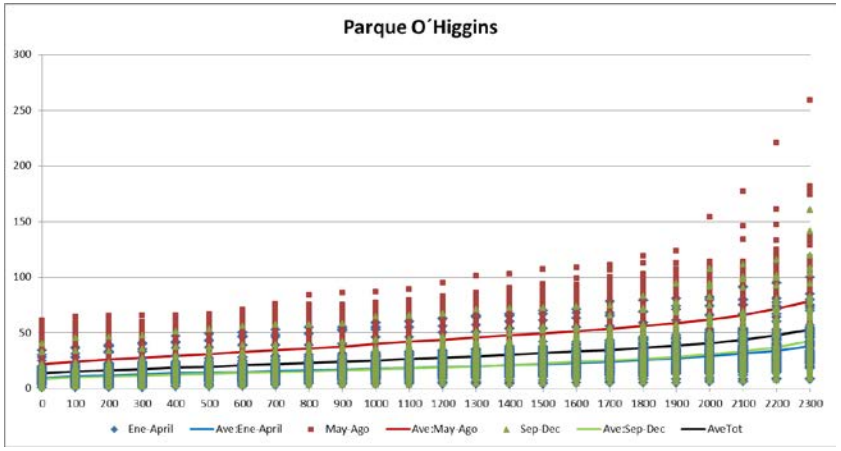
In order to see behavior throughout the day more clearly, all measurements for 0 AM, 1 AM, 2 AM, etc. were plotted for the whole year, as shown in Figure 4. The data was separated into the three periods of 4 months each mentioned before. Lines were added to show the hourly average for each period and the hourly average for the entire year (shown in black). Winter (May to August, in red) always has higher $PM_{2.5}$ concentrations than the annual average and the other two periods of the year. The periods of Jan-April (blue) and Sept-Dec (green) follow a similar pattern, with higher afternoon $PM_{2.5}$ concentrations in the Sept-Dec period than in the Jan-Apr period.

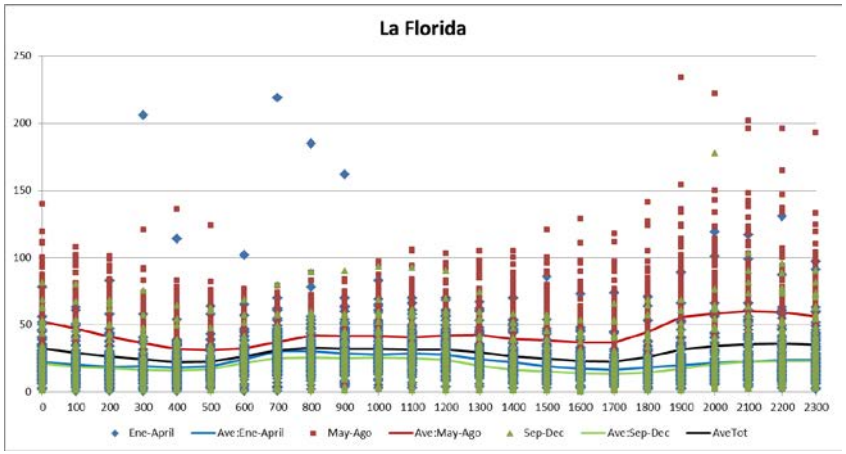
Stations show similar behaviors though the day except for the Parque O´Higgins station which is very different in that the $PM_{2.5}$ concentrations rise until 23:00 and then fall at 0:00. The reason for this has not been determined yet.

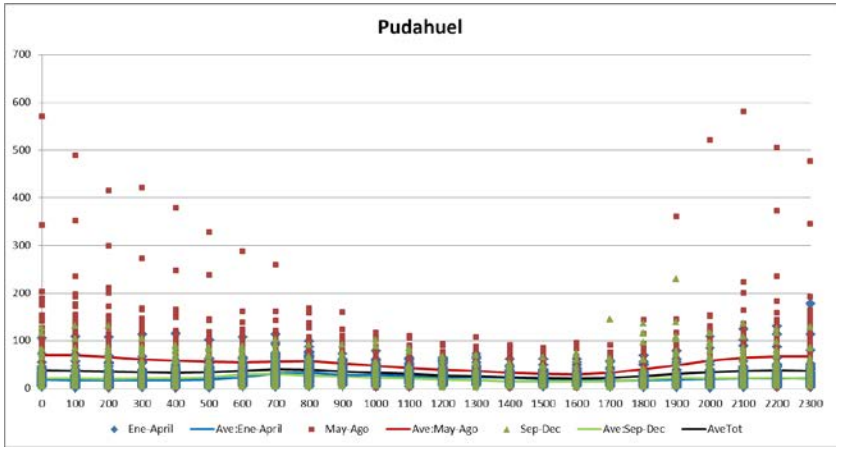
The Las Condes station registers the lowest concentrations of $PM_{2.5}$, which rise in the afternoon. The other stations (with the exception of Independencia) show that the concentrations are higher in the morning and afternoon, reaching their lowest point past midday (between 15:00 and 16:00).

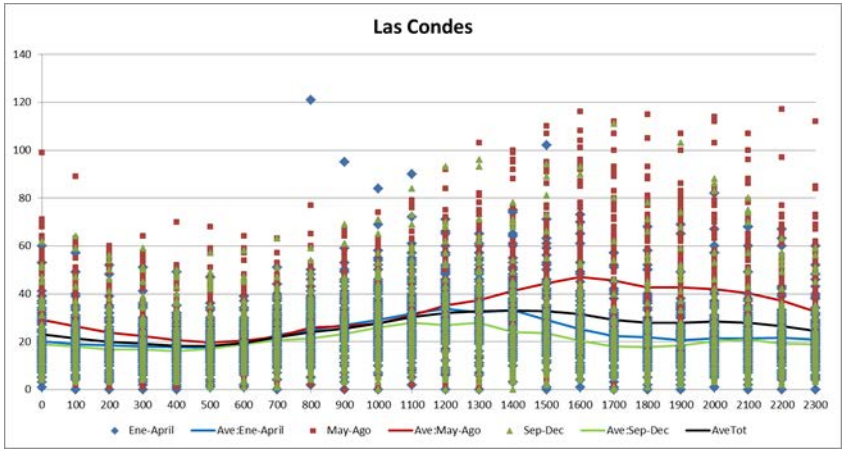
Figure 4: Hourly $PM_{2.5}$ concentrations for each station

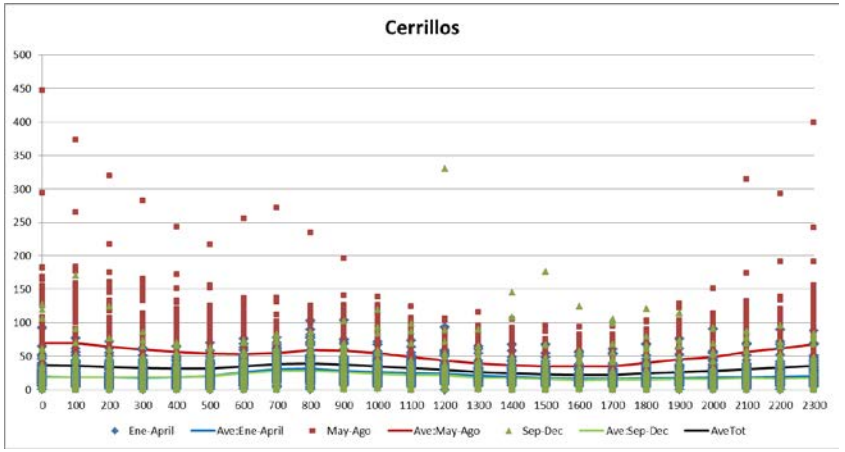


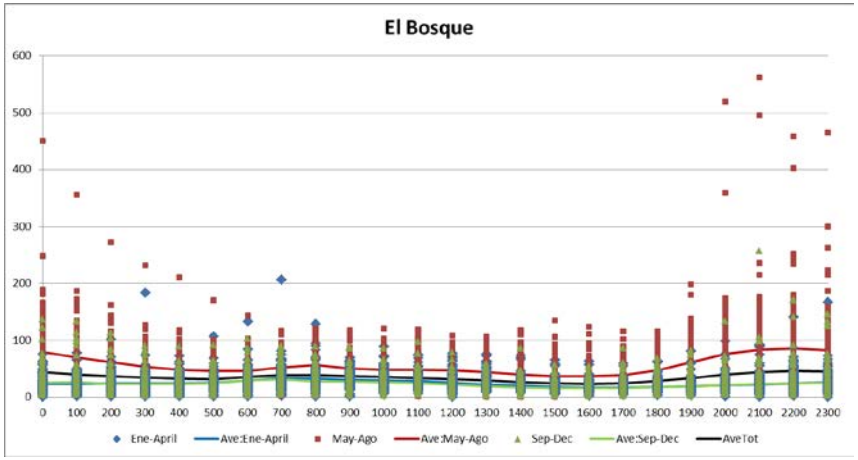


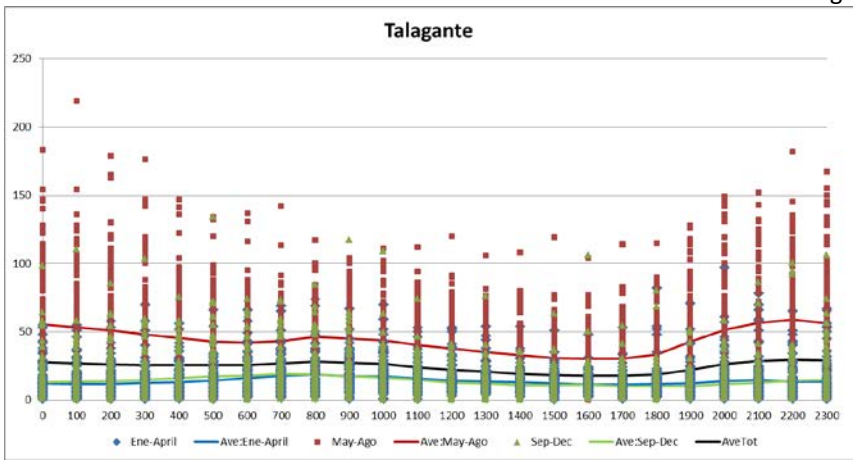


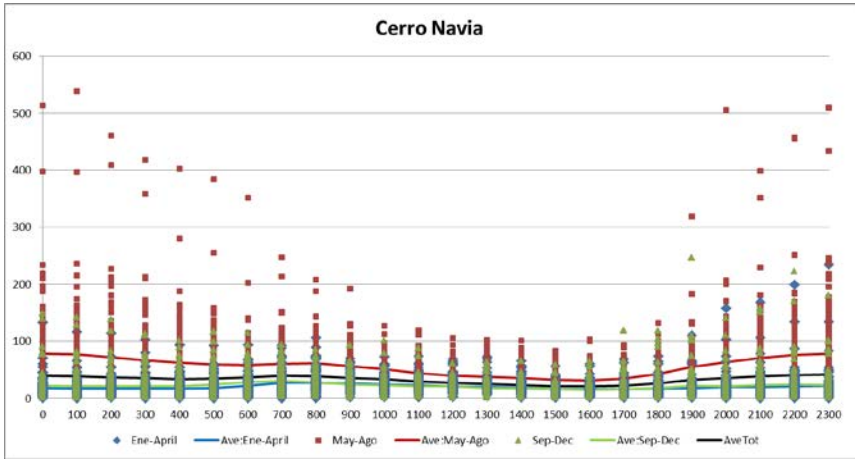


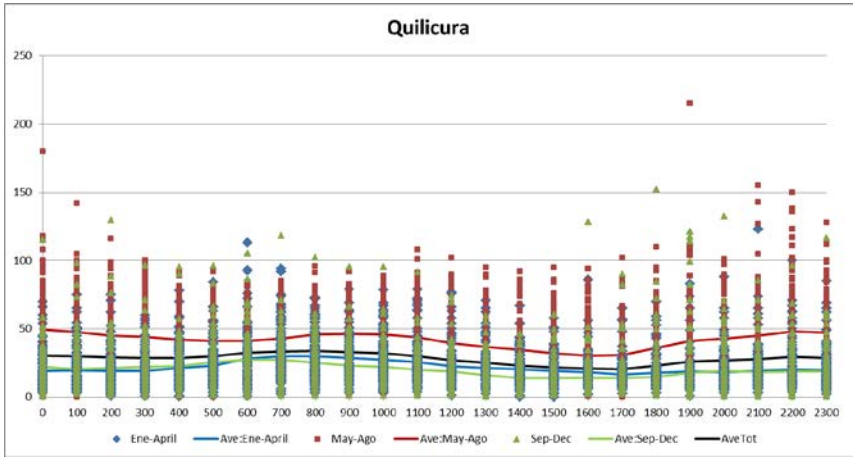


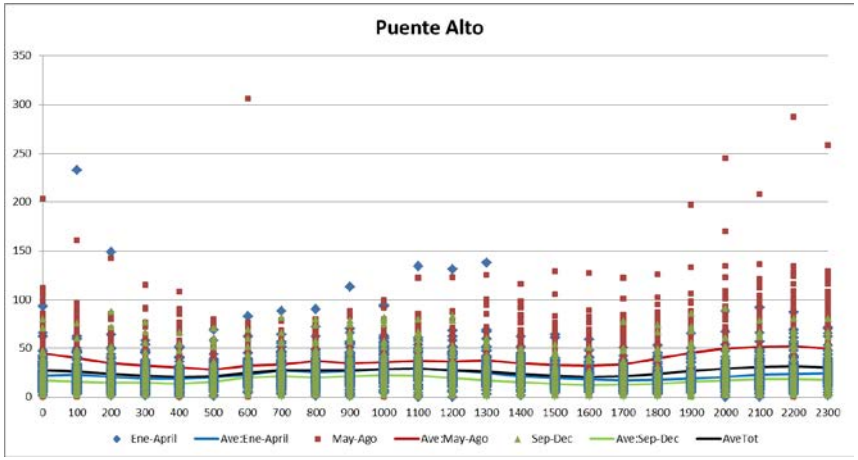



El Bosque
Ene-April
Ave:Ene-April
May-Ago
Ave:May-Ago
Sep-Dec
Ave:Sep-Dec
AveTot


Talagante
Ene-April
Ave:Ene-April
May-Ago
Ave:May-Ago
Sep-Dec
Ave:Sep-Dec
AveTot


Cerro Navia
Ene-April
Ave:Ene-April
May-Ago
Ave:May-Ago
Sep-Dec
Ave:Sep-Dec
AveTot


Quilicura
Ene-April
Ave:Ene-April
May-Ago
Ave:May-Ago
Sep-Dec
Ave:Sep-Dec
AveTot


Puente Alto
Ene-April
Ave:Ene-April
May-Ago
Ave:May-Ago
Sep-Dec
Ave:Sep-Dec
AveTot

To investigate the daily cycles further, a general relative semivariogram was calculated for each station based on formulas from Deutsch C. & Journel A., 1992 and Pitard F., 1993 applied to one dimensional datasets. These semivariograms allow a better assessment of cycles in $PM_{2.5}$ concentrations and the semivariogram nugget effect can be used to provide an estimation of the error associated with assigning a concentration to one hour, 24 hours or any other time period that may be of interest.

The traditional semivariogram is defined as follows:

Semivariogram:

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

The semivariogram value is defined as half of the average squared difference between two values of $PM_{2.5}$ separated by a distance h (lag).

General relative semivariogram:

$$\gamma_{GR}(h) = \frac{\gamma_{(h)}}{\left(\frac{m_{-h} + m_{+h}}{2}\right)^2}$$

The general relative semivariogram is the traditional semivariogram standardized by the squared mean of the $PM_{2.5}$ measurements used for each lag (h). This effectively normalizes the variance and allows direct comparison between stations.

The general relative semivariogram was calculated for the entire year for each station. Only the first 200 lags (there are about 1470 lags total) are shown in Figure 6. General relative semivariograms were also calculated for the three 4 month periods and they look very similar to the annual ones as shown in Figure 5.

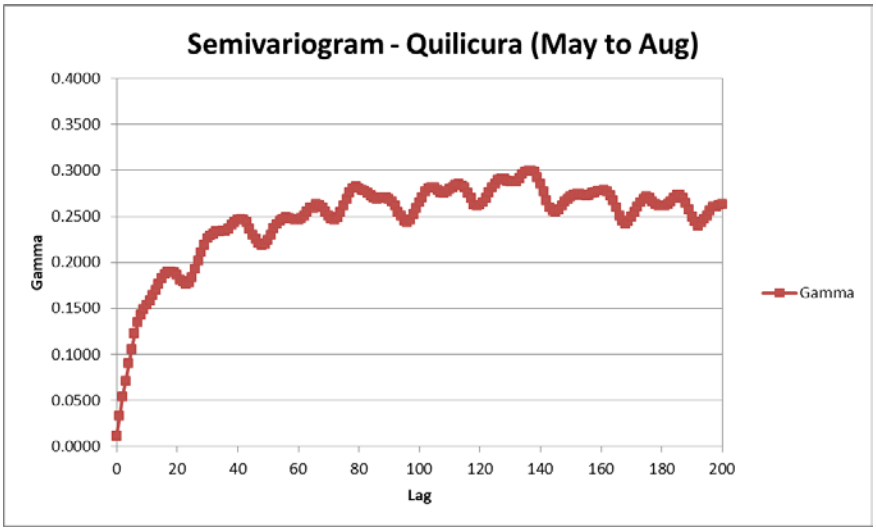


Figure 5: General Relative Semivariogram for $PM_{2.5}$ concentration from May to August, Quilicura station.

The graphs show a main cycle of 24 hours for all the stations (24 hours pass between one low point and the next in the semivariogram graphs) and a second cycle of 10 to 12 hours long for 6 of the stations (Independencia, La Florida, El Bosque, Puente Alto, Talagante and Quilicura). The secondary cycle is probably related to peak traffic hours (morning and evening rush hours, which in Santiago occur at 7 to 9 AM and at 7 to 9 PM).

The five stations with no secondary cycle are: Las Condes, Parque O´Higgins, Pudahuel, Cerrillos and Cerro Navia).

The Parque O´Higgins station shows a sharp drop of values between 23:00 and midnight as was detected in the time series graphs (Figure 4).

In the graphs, gamma (on the y-axis) gives an indication of the variability associated with each lag and represents the variance in $PM_{2.5}$ concentrations. Pudahuel, Cerro Navia and Talagante have the highest annual variability and Las Condes has the lowest variability, but shows more variability in winter that some of the other stations, which may be due to wood burning fireplaces used for heating (which are common in this area of the city and they are used in the afternoons and evenings when $PM_{2.5}$ concentrations rise in Las Condes). These differences are also shown in Table 3.

Table 3 shows relative variances (variance / average^2) for each station's data for the entire year and for the three 4-month periods. These relative variances correspond roughly to the sill (maximum value) of the semivariograms and are a measure of how variable $PM_{2.5}$ concentrations are at each station, with higher values indicating higher variability.

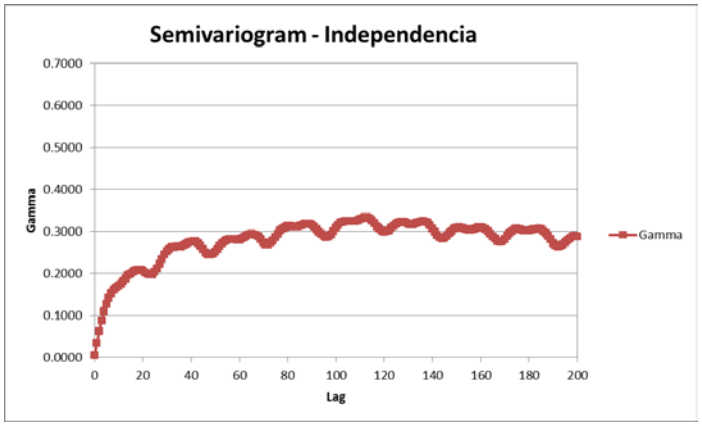
	D11	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	**D18**	**D27**	**D28**	**D29**
Relative Variance Jan-Dec	0.46	0.56	0.43	0.59	1.04	0.85	0.84	1.16	0.60	1.02	0.55
Relative Variance Jan-Apr	0.34	0.45	0.28	0.40	0.50	0.41	0.44	0.68	0.35	0.60	0.47
Relative Variance May-Aug	0.26	0.34	0.40	0.32	0.71	0.47	0.55	0.72	0.44	0.46	0.29
Relative Variance Sep-Dec	0.43	0.49	0.42	0.56	0.74	0.78	0.58	0.75	0.56	0.93	0.64

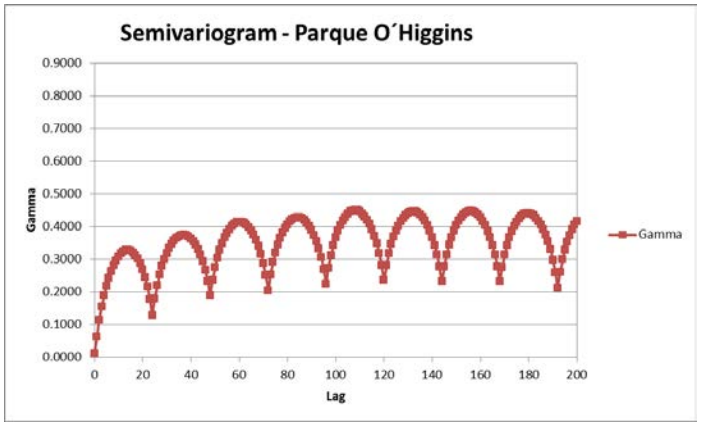
Table 3: General Relative Gamma values for $PM_{2.5}$ concentrations per station.

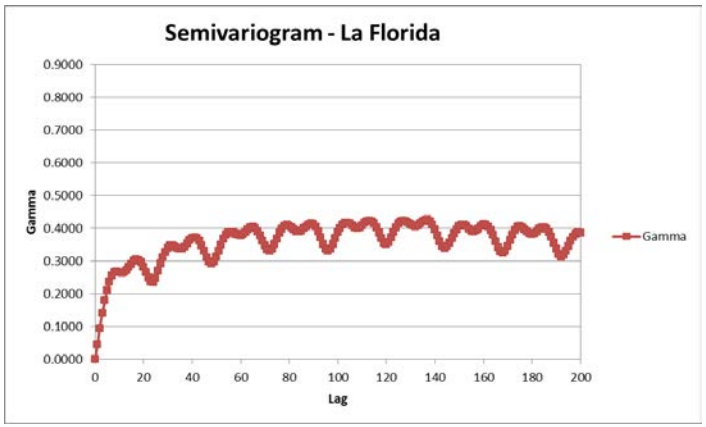
It is apparent that the Sep-Dec period has the highest variability (although not the highest average concentration) of $PM_{2.5}$ values for all stations.

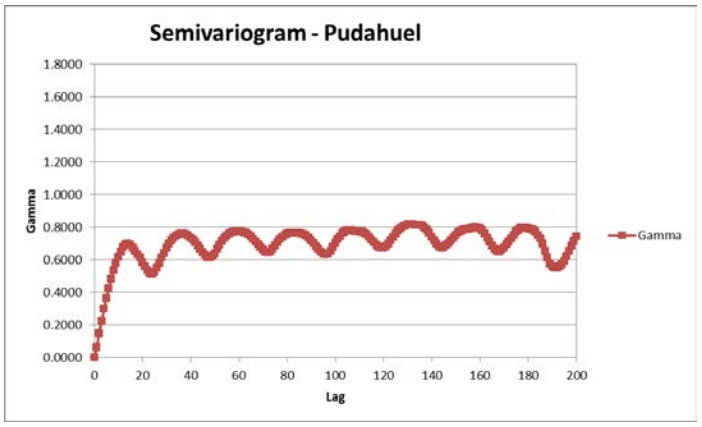
Several stations have lowest relative variances in the Jan-Apr period (Las Condes D13, Pudahuel D15, Cerrillos D16, El Bosque D17, Cerro Navia D18 and Puente Alto D27), while the others (Independencia D11, La Florida D12, Parque O'Higgins D14, Talagante D28 and Quilicura D29) have lowest relative variances in winter (May-Aug).

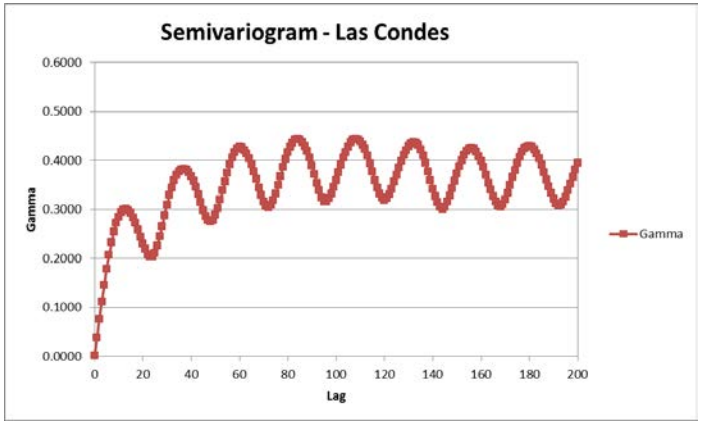
Figure 6: $PM_{2.5}$ concentration General Relative Semivariograms for each station.

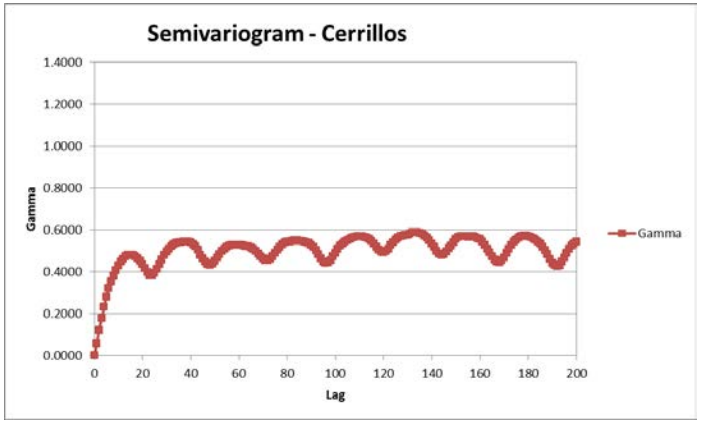


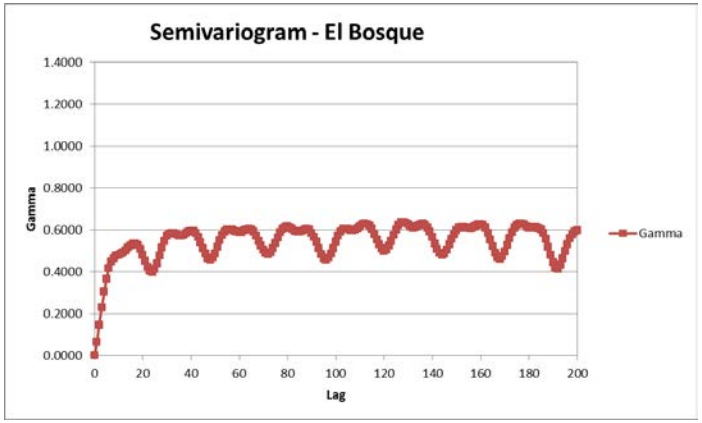


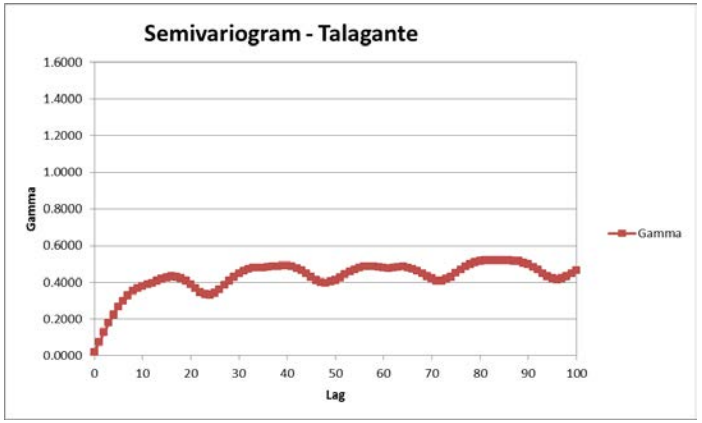


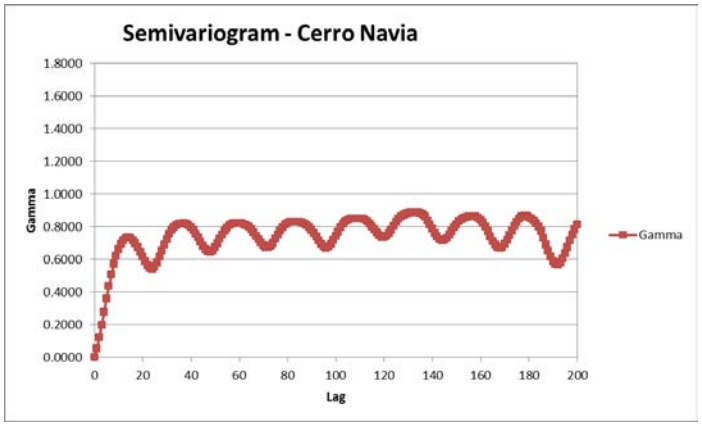


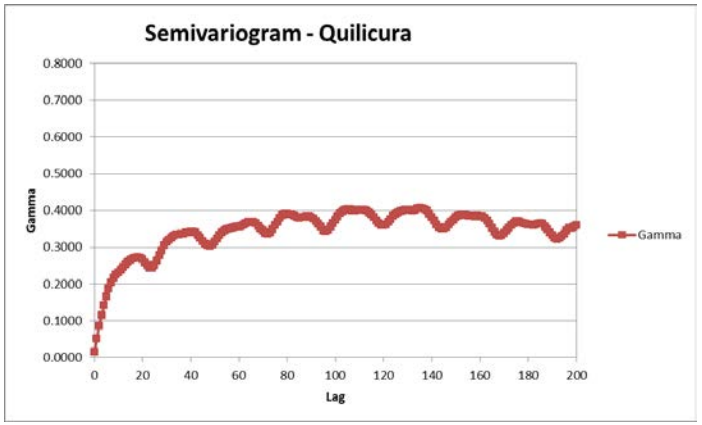


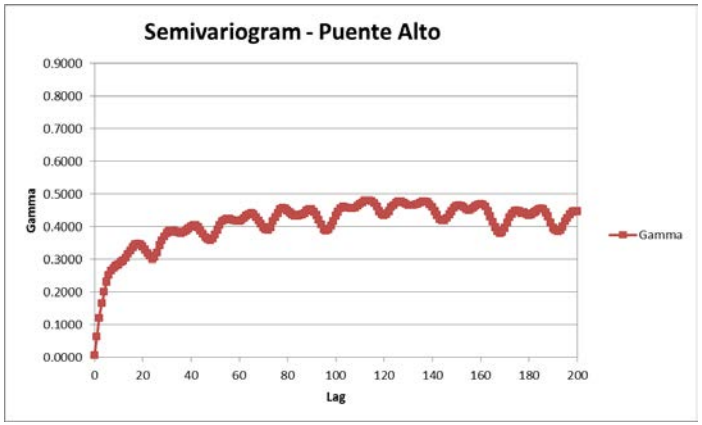



Semivariogram - El Bosque
Gamma
Lag
Gamma


Semivariogram - Talagante
Gamma
Lag
Gamma


Semivariogram - Cerro Navia
Gamma
Lag
Gamma


Semivariogram - Quilicura
Gamma
Lag
Gamma


Semivariogram - Puente Alto
Gamma
Lag
Gamma

Finally, the annual data was aggregated by day of the week and separated into the three 4-month periods to see how the concentration of $PM_{2.5}$ varies during the week. Figure 7 shows the average concentration of $PM_{2.5}$ by the day of the week.

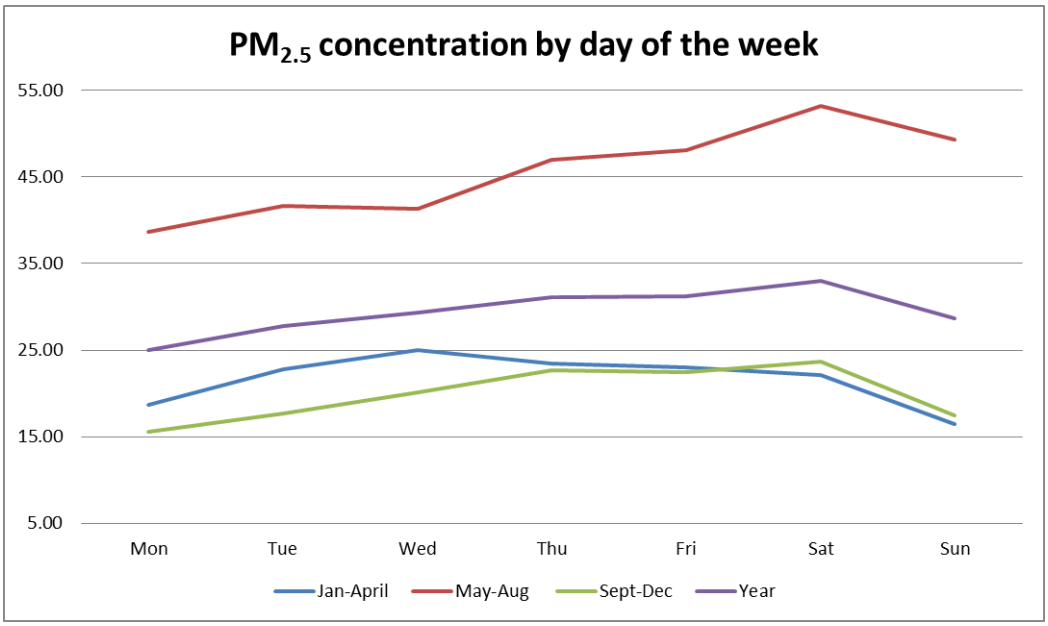


Figure 7: $PM_{2.5}$ concentration by day of the week

During colder months (May-Aug) and for the entire year (Jan-Dec), the highest concentration is reached over the weekend (Saturday), declining by Monday and slowly rising again towards the end of the week. During warmer months (Jan-Apr and Sep-Dec) the highest concentration is reached by mid-week (Wednesday or Thursday).

REGRESSION MODELS

In order to investigate if any of the environmental factors measured at the monitoring stations (wind speed and direction, relative humidity (RH), air temperature and elevation) have an effect on $PM_{2.5}$ concentrations, a regression analysis was performed for each station (Table 4) using the Excel data analysis regression function.

Table 4: Regression model per station

	D11	D12	D13	D14	D15	D16	D17	D18	D27	D28	D29
Adjusted R^2	0.11	0.12	0.06	0.08	0.12	0.16	0.16	0.20	0.12	0.26	0.15
Intercept	47.18*	48.84*	5.21*	66.82*	25.18*	54.82*	86.92*	102.44*	15.17*	88.18*	29.06*
Relative Humidity	-0.03*	-0.05*	0.21*	-0.28*	0.13*	0.05*	-0.19*	-0.29*	0.26*	-0.33*	0.04*
Temperature	-0.63*	-0.77*	0.23*	-1.42*	-0.23*	-0.69*	-1.83*	-2.37*	0.16*	-2.09*	-0.19*
Wind Speed	-5.89*	-6.88*	0.17*	-1.64*	-3.62*	-3.02*	-2.55*	-3.40*	-3.39*	2.36*	-3.69*
Wind Direction	-.004*	-0.01*	0.04*	0.02*	0.023*	-0.05*	-0.04*	-0.04*	-.0005*	-0.06*	0.013*

* Significant p-value at 95%

Regression models constructed for each station have very low adjusted R^2 values, and high intercepts (varying between 5.21 μg/m^3 for Las Condes - D13 and 102.44 μg/m^3 for Cerro Navia - D18). In general, temperature and wind speed have large negative coefficients, except for 2 stations. Increases in wind speed increase Pm2.5 concentrations for Las Condes and Talagante, which are located in opposite directions from the center of Santiago and receive air pollution from the city depending on wind direction. Increases in temperature decrease $PM_{2.5}$ concentrations, except at the Las Condes and Puente Alto stations, which have the highest elevations. Relative humidity has coefficients that vary in magnitude and sign. As relative humidity rises in winter, some stations located on the western side of Santiago (O'Higgins, El Bosque, Cerro Navia and Talagante) experience lower levels of $PM_{2.5}$. These stations also report lower levels of $PM_{2.5}$ when the air temperature rises. Wind direction has very low coefficients and varying signs.

Another regression model was performed for all stations together (Table 5) using the Ordinary Least Squares tool in ArcMap. This model also included for each station's elevation, distance to main roads, the number of point sources of $PM_{2.5}$ pollution within a 5 km radius and the amount of pollution produced by those sources (kg per day). The total amount of records with no missing values for the dependent or independent variables was 75,315.

Table 5: Regression model for all stations together

Variable	Coefficient	Range of values
Adjusted R^2	0.113	
Intercept	39.47*	
Relative Humidity	-0.004	2.7 to 105 %
Temperature	-1.001*	-5 to 37 ºC
Wind Speed	-1.62*	0 to 20 m/sec
Wind Direction	-0.003*	0 to 359 º
Elevation	0.008*	309 to 789 masl
Number of pollution sources	0.014*	4 to 398
Pollution per day	-0.009*	5.5 to 509 kg/day
Distance to main roads	0.005*	173 to 2295 mts.

* Significant p-value < 0.01

The results were very weak, but the coefficients for each independent variable were similar to the ones reported in the literature review. The regression model's intercept (of 39.47 $\mu g/m^3$) is high relative to Chile and WHO limits, but is reduced by increases in environmental variables (relative humidity, temperature, wind speed and wind direction). Of these, temperature and wind speed have the largest coefficients and, considering the range of possible values for each variable, will have the largest effect on $PM_{2.5}$ concentrations. Elevation has a very small positive coefficient, perhaps as a result of heating in winter. As can be expected, the number of surrounding pollution sources and distance to main roads both have positive coefficients. Pollution per day, however, has a negative coefficient, perhaps resulting from one-time voluntary estimates from several years ago which do not necessarily correspond to current conditions.

Although the histogram of the residuals for each regression model appeared to follow a normal distribution, the Jarque-Bera statistic was significant at 0.01, indicating that the model predictions are biased. Also the graph of predicted vs residuals did not look random, which confirms this result.

The regression models are not good models for explaining or predicting the variability observed in the concentration of $PM_{2.5}$ through the year. Since $PM_{2.5}$ concentrations change throughout the year the mean and variance of the data are not constant, making the regression models unsuitable.

THREE-DIMENSIONAL KRIGING

In order to interpolate $PM_{2.5}$ concentrations across the city and through time, three dimensional Kriging was performed using GSLib software. The time stamp for each hourly measurement was used as the Z coordinate.

The first step was to organize the data in the format that is read by GSLib programs. X and Y values were defined by the location of each measurement station, projected to WGS 1984 UTM Zone 19S.

The time column (Z) which contains date and time data in the format of YY/MM/DD HH:MM was modified to Z values with increments of 1. Each station was assigned Z values from 1 to 8784 to represent the hourly measurements taken throughout the year. A Z coordinate of 1 was assigned to measurements taken on January 1, 2016 at midnight, 2 for measurements taken on January 1, 2016 at 1:00AM and so forth, ending at Z coordinate of 8784 for measurements taken on December 31, 2016 at 11:00PM[1].

Two experimental semivariograms using the data for the whole year (i.e. values calculated from the measured data) were calculated:

- The <u>omni-directional horizontal</u> semivariogram considered only one period of time (1 hour). A band width of 5000 meters, a lag size of 2000 meters and 10 lags with a lag tolerance of 1000 meters were used (Figure 8).

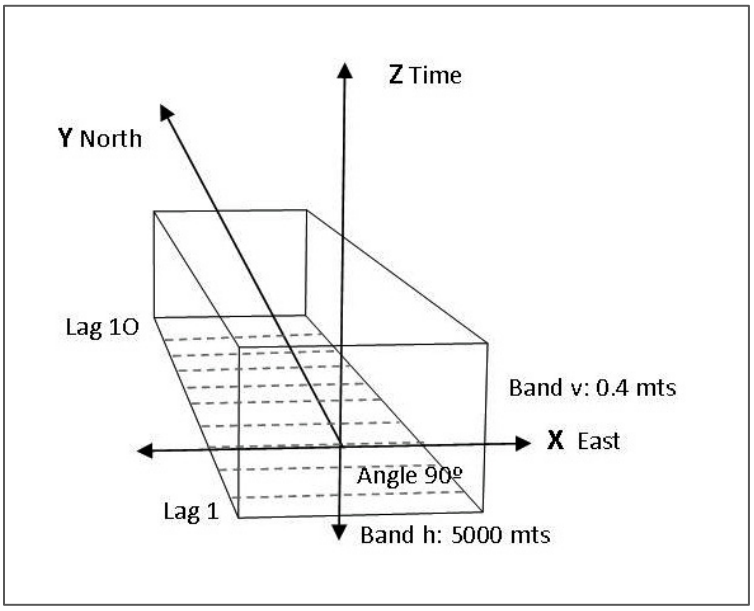


Figure 8: Omni directional horizontal semivariogram illustration.

[1] The Year 2016 was a leap year.

- The vertical semivariogram was set up to use values from one station at a time by defining a lag size of 1, 168 lags (one week of values) and a lag tolerance of 0.5. The vertical bins were allowed to extend 1000 m out from each station in the East, West, North and South directions. Since the closest stations are at least 2000 meters apart, this setup ensured that pairs would not include data points from two or more stations, creating a purely vertical semivariogram (Figure 9).

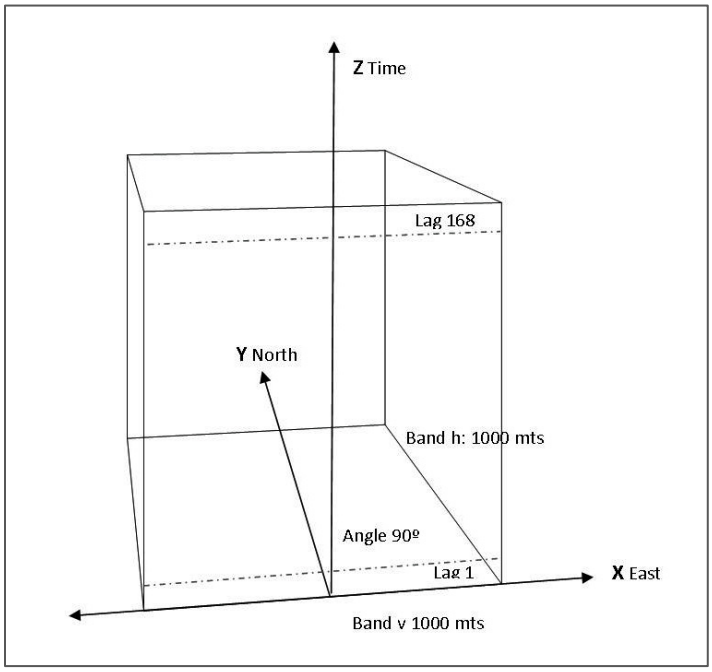


Figure 9: Vertical semivariogram illustration.

The experimental semivariogram values are shown in figures 10 and 11 as red points (joined by red lines). Models were fitted to the experimental data and are shown in blue points (joined by blue lines) in the same figures.

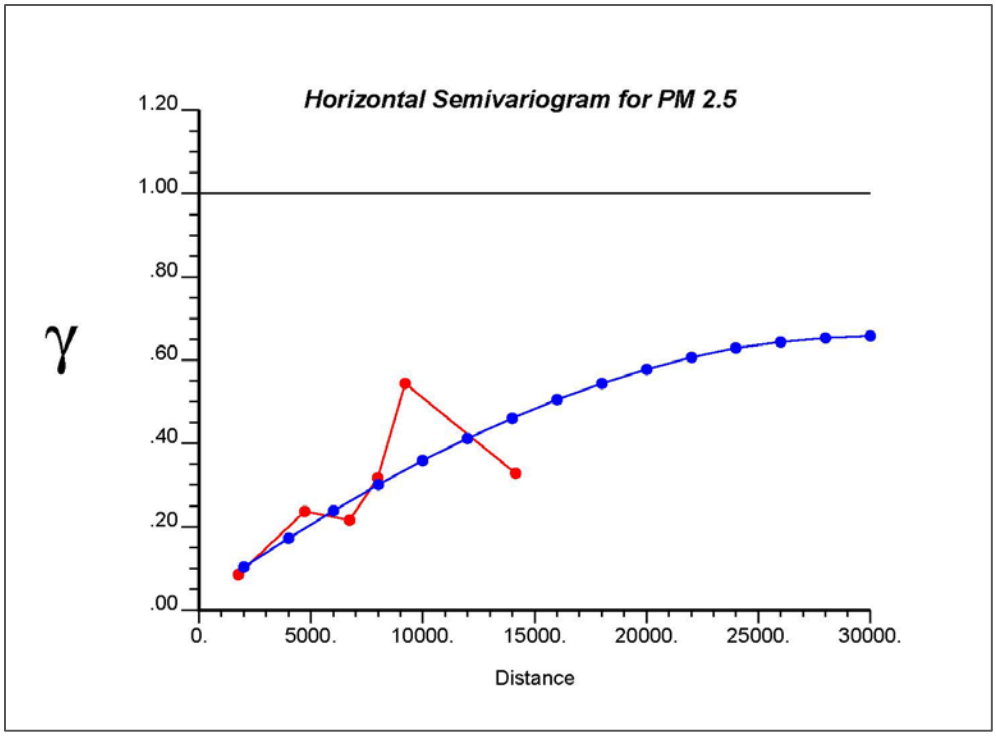


Figure 10: Horizontal Semivariogram´s experimental and fitted model

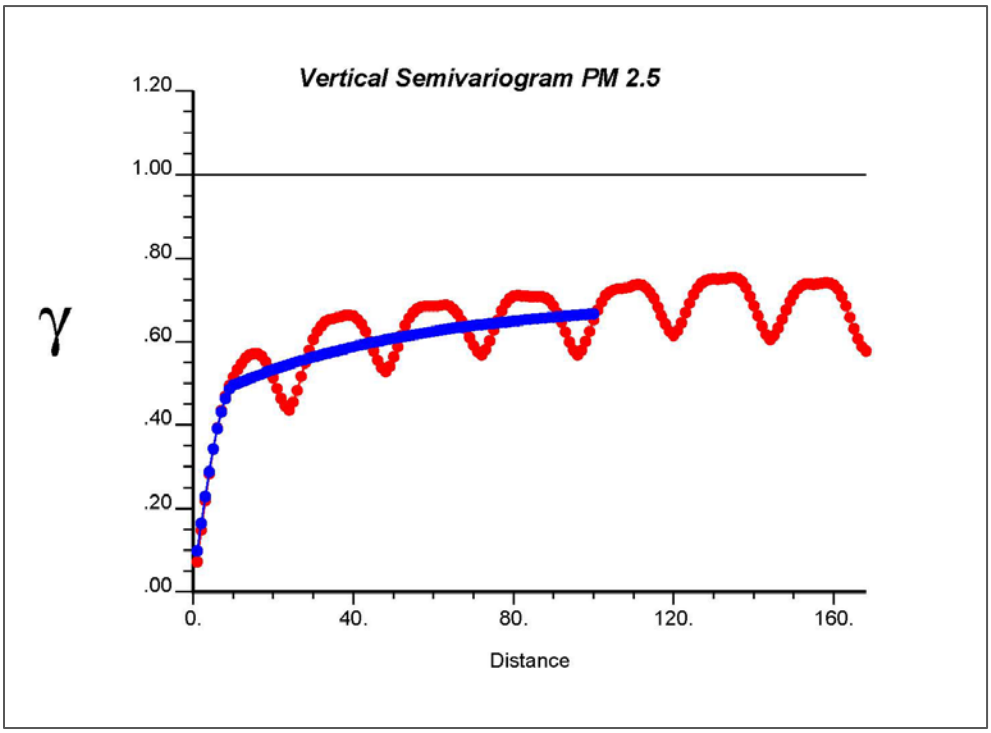


Figure 11: Vertical Semivariogram´s experimental and fitted model

The omnidirectional horizontal semivariogram model is a composite model with a nugget effect of 0.03, a spherical model with a range of 28000 m and variance contribution of 0.42, followed by an exponential model with a range of 50000 m and variance contribution of 0.25. No anisotropy was included because the $PM_{2.5}$ concentrations vary throughout the year and anisotropy may vary from one time step to another.

The vertical semivariogram (though time) was modeled using the same composite model as described above, but the spherical model had a range of 10, which corresponds to 10 time steps of 1 hour each (shown in Figure 10). The range of 10 indicates that $PM_{2.5}$ concentrations are correlated up to 10 time steps (hours). Beyond 10 hours, correlation decreases and the cyclical pattern in $PM_{2.5}$ concentrations is apparent (as found in the general relative semivariograms constructed for each station). The cyclical component was simplified and modeled as slowly increasing exponential model with a range of 150 hours. The nugget effect and variance contributions are the same as those used for the horizontal model (since both models must be combined in order to perform 3D kriging).

Once the semivariograms models had been fitted to the experimental data, Ordinary 3D Kriging was run to create a grid of interpolated values in space and time. Ordinary kriging was chosen because it assumes that a constant mean is unknown, $Z(s) = \mu + \varepsilon(s)$[2] and is capable of replicating trends in the data (in this case both in space and time). Kriging was set up to produce an estimated $PM_{2.5}$ concentration at 500 m intervals (in the North and Easterly directions over a large enough extent to cover the city) and every hour. The search radii used were 28000 m in the horizontal and 150 hours in the vertical directions. Due to limitations of the software, kriging was performed per month, producing almost 6,000,000 interpolated values per month.

After kriging, each time surface (slice) was extracted to create 8784 layers that were displayed in an animated video that shows how the $PM_{2.5}$ concentrations change over time for the study area (see the attached mp4 video files).

The thresholds for pollution levers in Santiago are as follows:

Level	**µg/m^3**
Good	Less than 50
Regular	50-79
Alert	80-109
Pre-emergency	110-169
Emergency	Above 170

The videos are color coded from 10 to 170 µg/m^3, with low values depicted in blue and high values shown in red. $PM_{2.5}$ concentrations lower than 10 µg/m^3 are also shown in blue, and concentrations over 170 µg/m^3 are also shown in red.

[2] Ordinary kriging has an advantage over Simple kriging in that it replicates spatial trends by kriging using a local rather than a global (fixed) mean.

As a reference, The World Health Organization defines the guidelines for $PM_{2.5}$ as a maximum of 10 μg/m^3 for the annual average and a limit of 25 μg/m^3 for the 24-hour mean (not to be exceeded for more than 3 days/year). These levels are clearly exceeded in Santiago.

June is a good example of the cyclic nature of the $PM_{2.5}$ concentrations. There were two events (June 18 and 27) where emergency measures were imposed due to the high levels of pollution in Santiago. January shows the same cyclic variation but at a much lower pollution level, although it also has an extraordinary event which occurred when a landfill to the south of Santiago caught fire and burned for several days, creating a sanitary emergency in the city due to pollution

EMERGING HOT SPOT ANALYSIS

The Emerging Hot Spot Analysis tool is part of the Space Time Pattern Mining toolbox. This toolbox has statistical tools for analyzing data distributions and patterns in the context of both space and time[3].

Interpolated surfaces were created for each month using kriging. Then X, Y and Z (representing a time stamp) coordinates were added to each interpolated point using GSLIB software. The files were then transformed to text files to be able to import them as tables within ArcGIS. An XY event layer was created for each table and exported as a point feature to a file geodatabase. A python script was developed to add the time stamp (day, month, year and hour in the correct format for the Create Space Time Cube tool) to each point feature based on its Z value. Finally, the merge tool was used to create a point feature for the whole year (70, 087,536 points in total, 5.7 GB file geodatabase)

This point feature was used to create a Space Time Cube (using the Create Space Time Cube tool) which is necessary step prior to running the Emerging Hot Spot Analysis tool. The space time cube is a netCDF file which aggregates timestamped point features into space-time bins. Within each bin the points are counted and summary statistics are calculated (the options available are: sum, mean, min, max, standard deviation and median)[4].

[3] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/an-overview-of-the-space-time-pattern-mining-toolbox.htm
[4] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm

A time cube illustration is shown in figure 12:

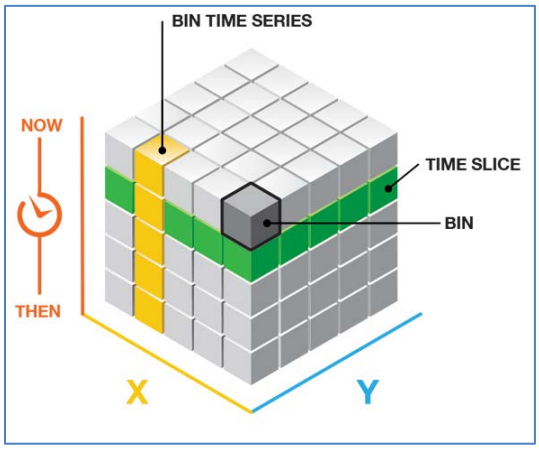


Figure 12: Illustration of a time bin cube[5] (source: ArcGIS online help)

The space time cube's parameters replicated the data interpolated by kriging since the time steps were set to 1 hour and the X Y distance interval was set to 500 meters. Because the data was evenly spaced by design, the start time was chosen for the time step alignment and the mean of the $PM_{2.5}$ points within each bin was selected as the summary statistic, which in this case corresponds to the interpolated $PM_{2.5}$ value since there would be only one point in each space-time bin.

The space time cube created had a point count of 7,979 locations over 8,784 time step intervals. Each bin is 500 meters by 500 meters spanning an area 50,500 meters west to east and 39,500 meters north to south. Each of the time step intervals is 1 hour in duration so the entire time period covered by the space time cube is 8,784 hours (which corresponds to the entire year of 2016). Of the 7,979 total locations, all of them contain one point for each time step interval. These 7,979 locations cover all 70,087,536 space-time bin with no bins having missing values (because kriging was performed to generate values at the same spatial and temporal resolution). The overall trend of $PM_{2.5}$ was reported as decreasing (this is the time trend in $PM_{2.5}$ concentrations over the entire year).

A visualization of the values in the time cube is shown in figure 13. Due to memory limitations (the cube had a size of 3.6 GB) visualization of the cube was done using a second cube of 500 x 500 meters and 24 hour intervals (which only weighs 195MB and can be rendered on screen). This second cube contains 24 hour averages of $PM_{2.5}$ concentrations, but is detailed enough to view spatial and temporal trends.

[5] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/create-space-time-cube.htm

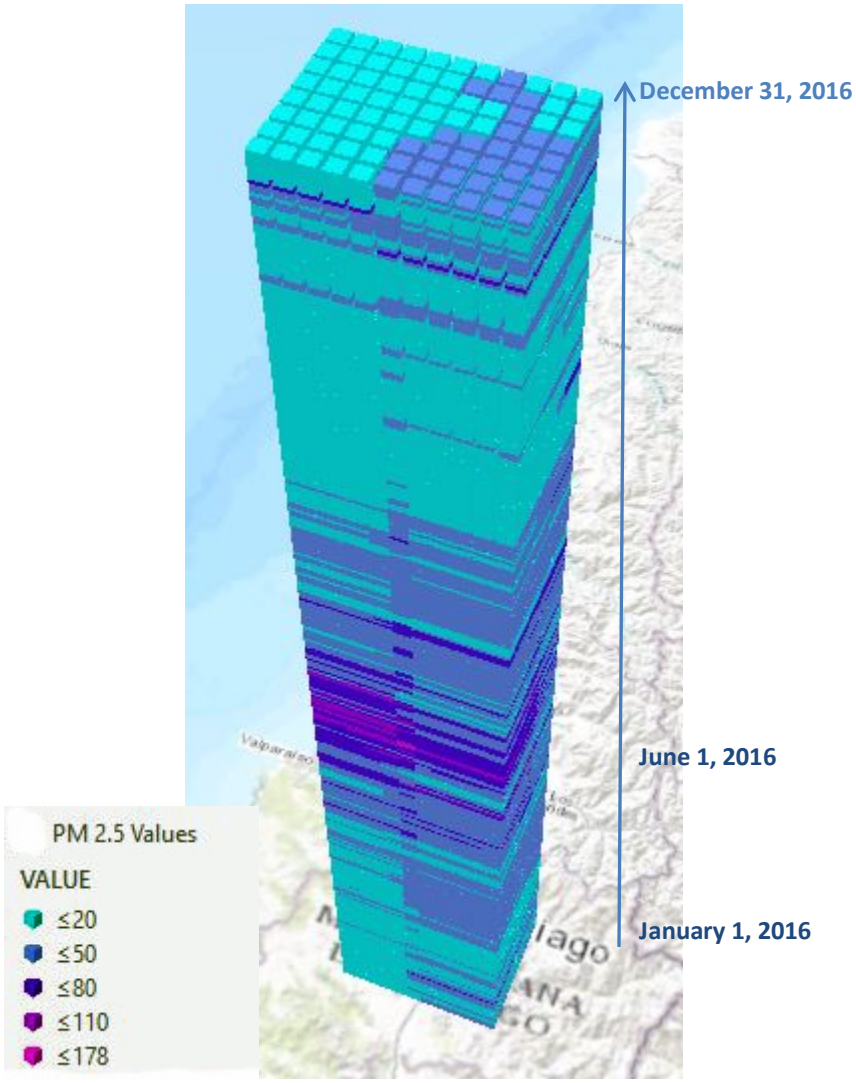


Figure 13: Space time cube visualization of $PM_{2.5}$ values in in 3D.

The Emerging Hot Spot Analysis tool identifies statistically significant trends in the clustering of point densities (counts) or summary fields (values of the variables) in the space time cube. It identifies if high or low values cluster in space and time assigning categories such as new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating, and historical hot and cold spots[6] to the results.

An Oscillating Hot Spot is defined as *"A statistically significant hot spot for the final time-step interval that has a history of also being a statistically significant cold spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots".*

[6] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/emerginghotspots.htm

An Oscillating Cold Spot is defined as *"A statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant cold spots".*[7]

The Emerging Hot Spot analysis tool was able to identify Oscillating Hot and Oscillating Cold spots for the whole year. The results are shown in figure 14.

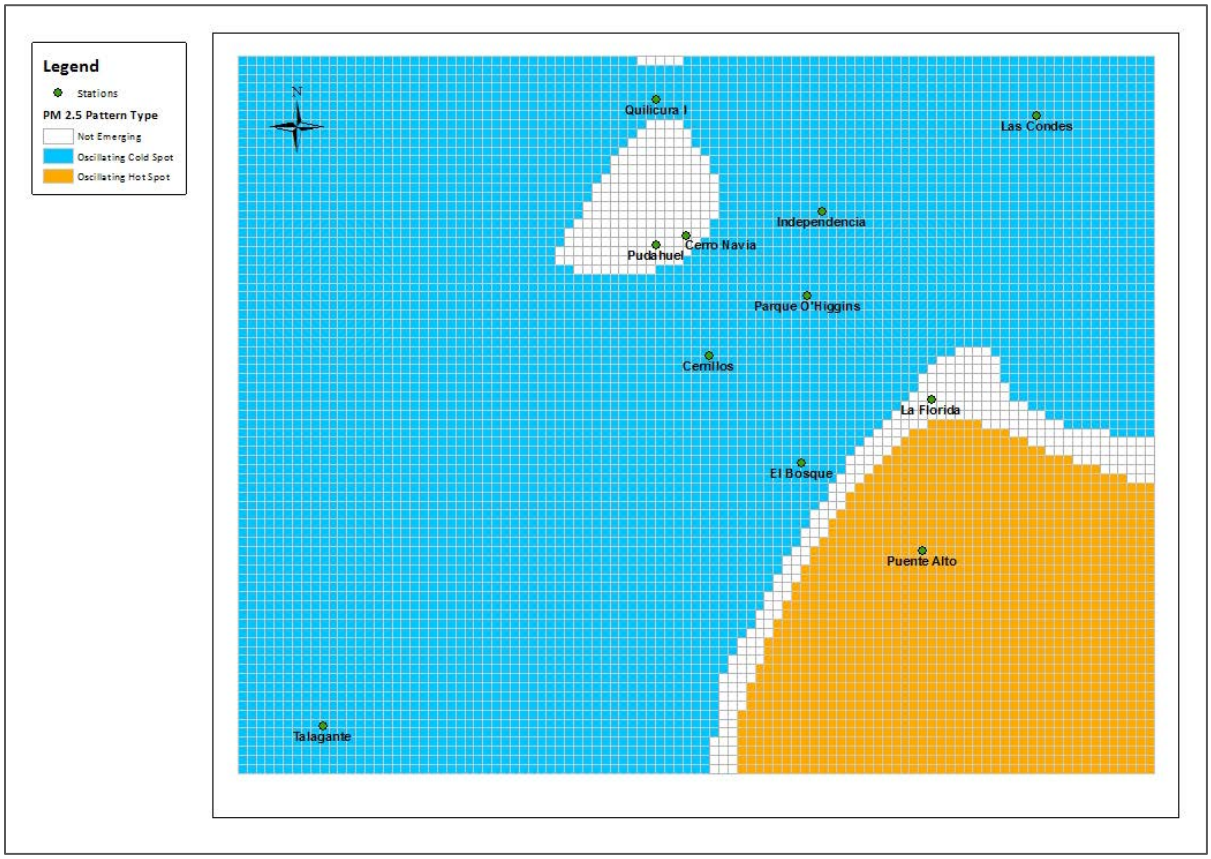


Figure 14: Emerging Hot Spot Patterns for $PM_{2.5}$.

Other space time cubes with larger cell sizes or time steps were tested, but the results were very similar. Results other than oscillating hot and cold spots are somewhat difficult to identify because the data is highly variable through time (hourly, daily and seasonally). Additionally, the categories of the hot spot analysis are based on behavior in the last time step of the cube relative to previous ones. This makes it hard for the tool to identify hot spots in cyclical data such as $PM_{2.5}$ concentrations (which are low in summer – at beginning and end of the year - and high in winter – midyear months, add to the fact that $PM_{2.5}$ has a daily and weekly cycle variations).

[7] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/learnmoreemerging.htm#GUID-09587AFC-F5EC-4AEB-BE8F-0E0A26AB9230

The Emerging Hot Spot tool adds its results to the space time cube, so variability through time can be seen when the space time cube is rendered to show hot/cold spots, as depicted in figure 15. In this case, as for figure 15, the visualization was done using a second cube of 500 x 500 meters and 24 hour intervals.

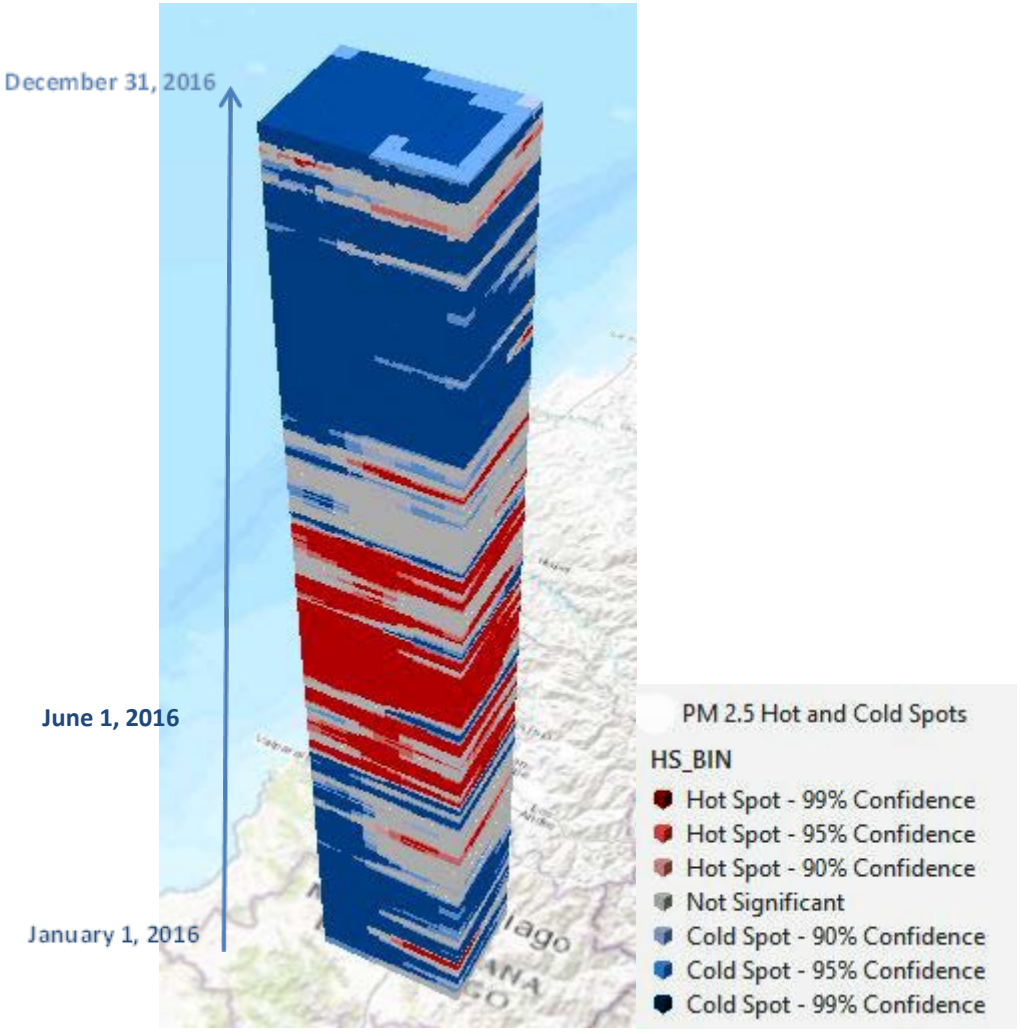


Figure 15: Hot and Cold Spots in the space time cube for $PM_{2.5}$ in in 3D.

The Emerging Hot Spot tool also shows the trend for each bin (downward, upward or not significant), and the percent of time steps in which the bin was classified as a significant hot or cold spot.

Figure 16 shows the trend for each bin over time. The majority of the bins had a downward trend (at a 99% confidence level) because the year starts with low values of $PM_{2.5}$ which increase in the winter month and decrease towards the end of the year. There is an area surrounding Cerro Navia, Pudahuel and Parque O'Higgins stations where no significant trend was detected.

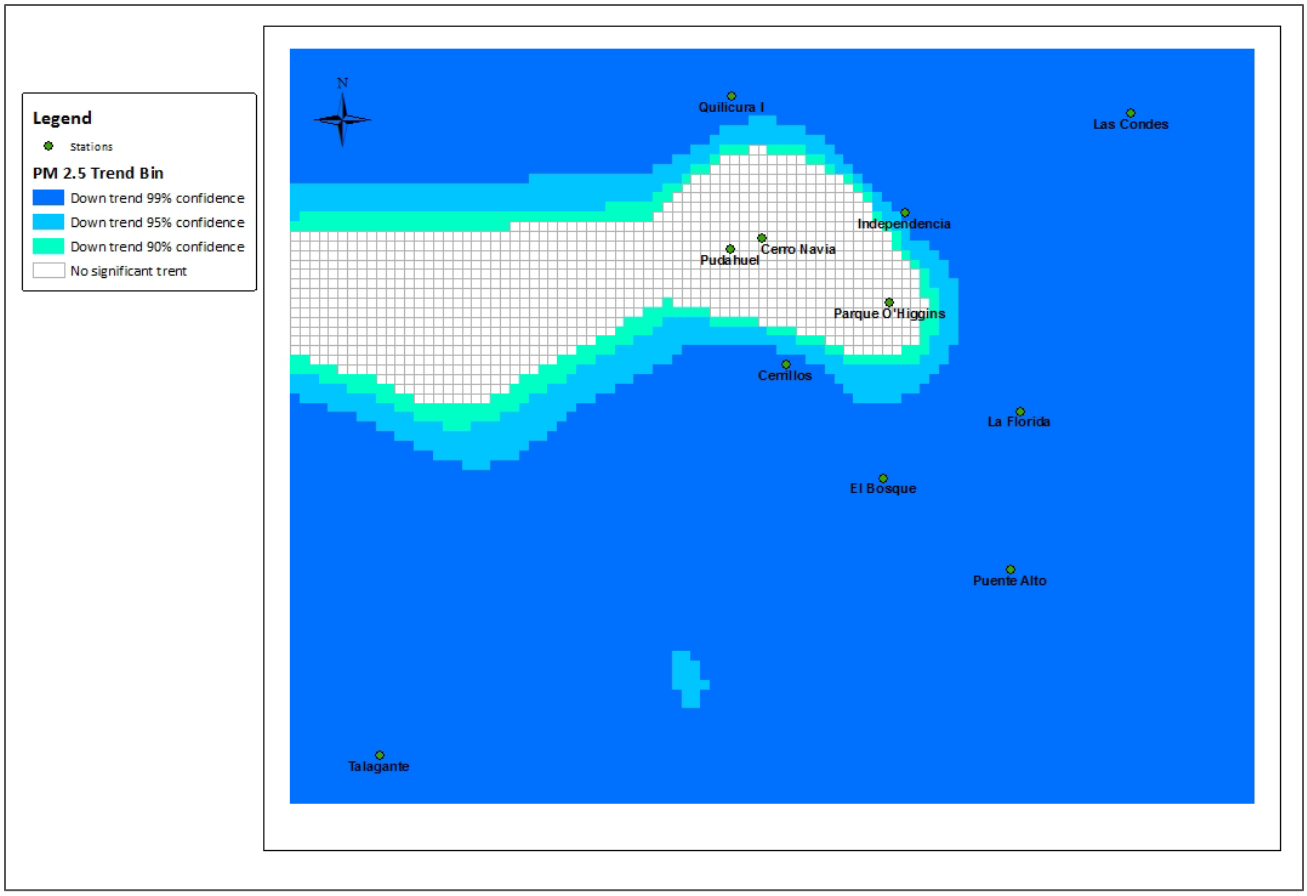


Figure 16: Emerging Hot Spot Trend for $PM_{2.5}$.

Figure 17 shows the percent of time that the bins were classified as statistically significant hot spots. The map indicates that all bins were classified as hot spots between 26 and 44 percent of the time (high concentration of $PM_{2.5}$ surrounded by bins with high $PM_{2.5}$ concentrations). The area surrounding the El Bosque station has the highest percentages of hot spot classifications.

The mean concentrations of $PM_{2.5}$ for each bin's time series vary between 23.8 and 34.1 $\mu g/m^3$, with an overall annual mean for the whole cube of 28 $\mu g/m^3$. From a health perspective it is interesting to examine the maximum values of $PM_{2.5}$ in each bin, which vary between 107.5 and 525.3 with an overall annual mean of 240.4, indicating that during some hours of the year Santiago's population is exposed to very high levels of $PM_{2.5}$.

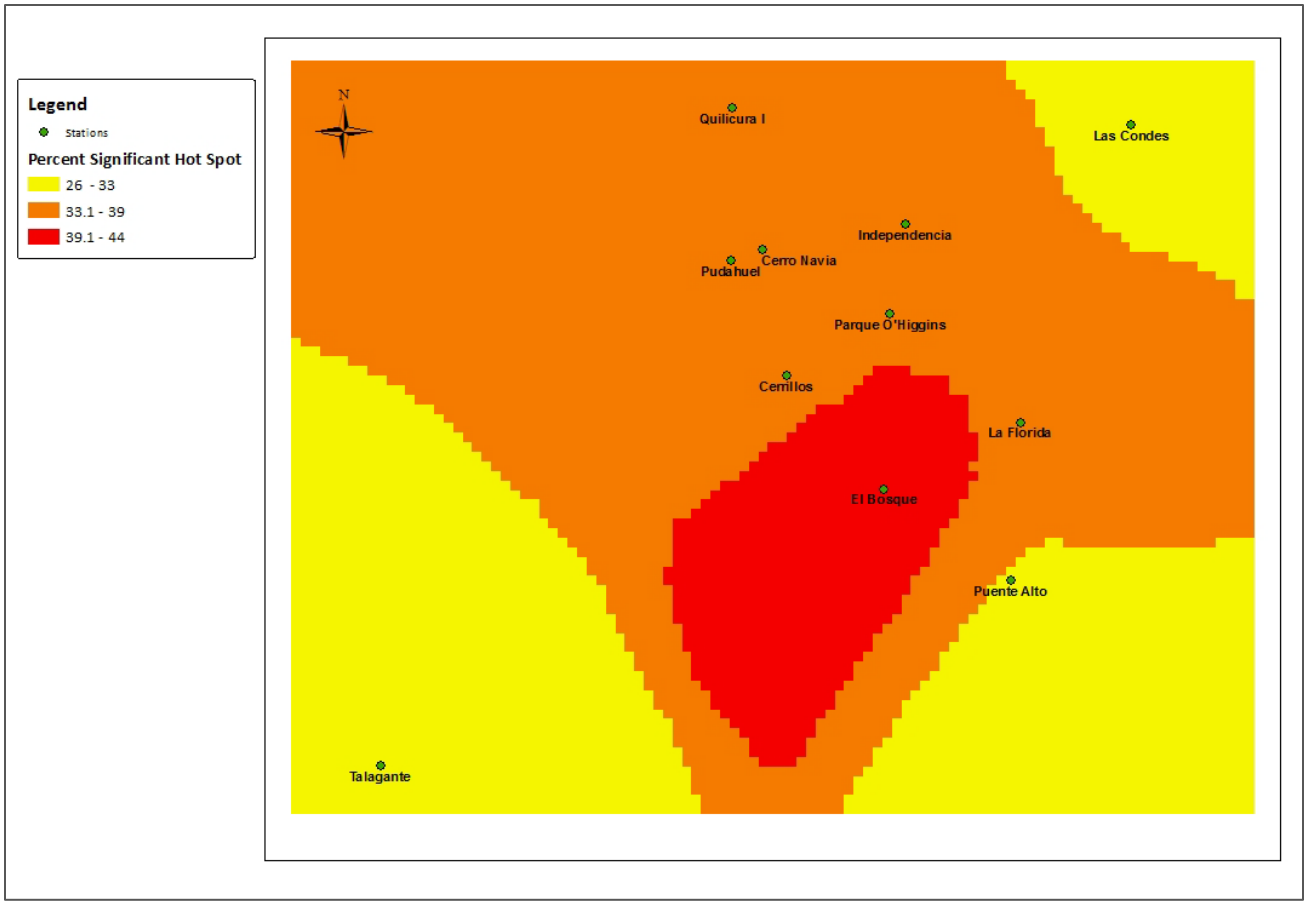


Figure 17: Percent of Hot Spot for $PM_{2.5}$ concentrations.

Figure 18 shows the percent of statistically significant cold spots over time in the study area. This map is an inverse of the hot spot areas, but also shows the area surrounding Las Condes, Puente Alto and Talagante stations as significant cold spots a high percentage of the year (between 63 and 70% of the time).

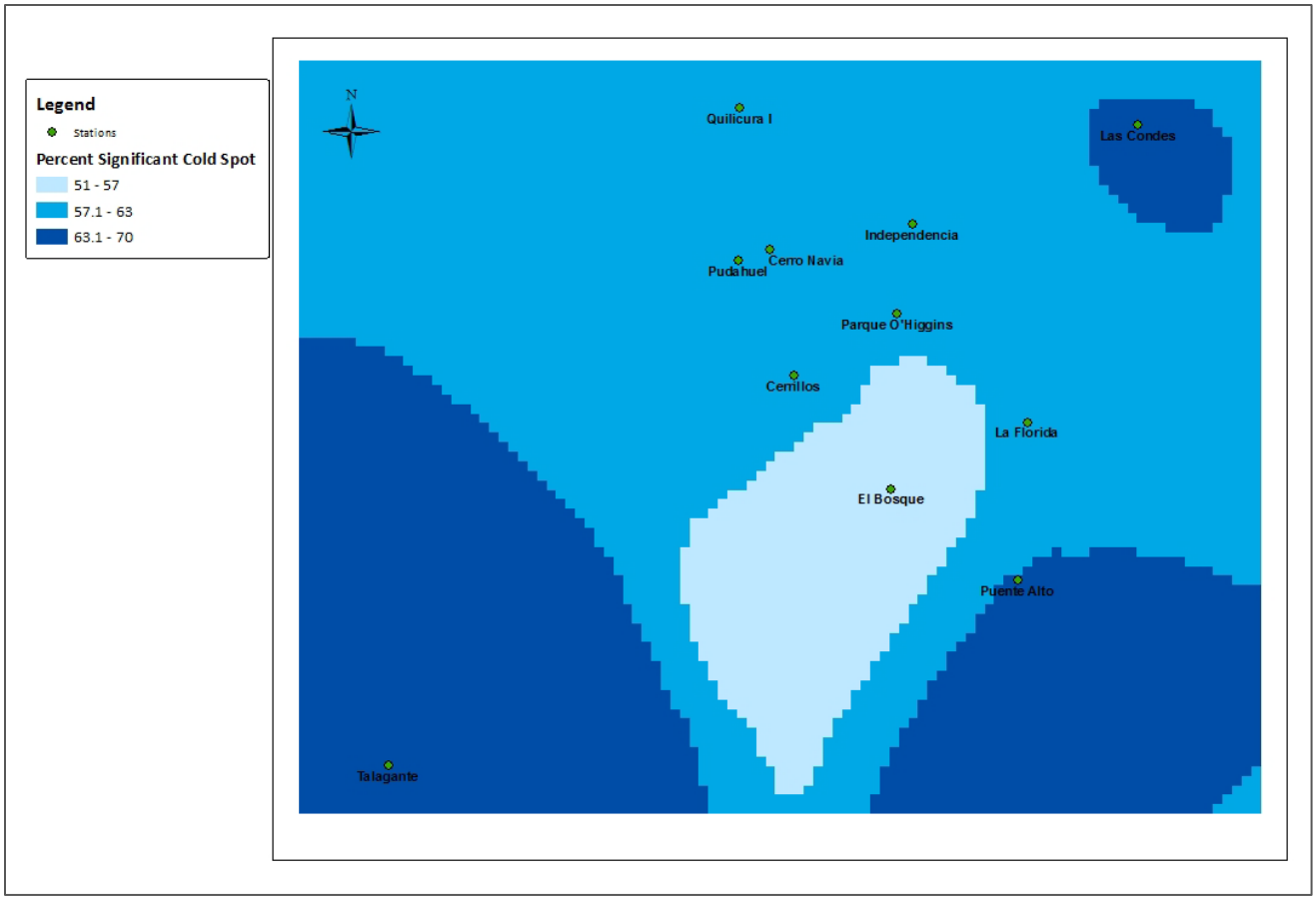


Figure 18: Percent of Cold Spot for $PM_{2.5}$ concentrations.

The Space Time Pattern Mining toolbox also includes a Local Outlier Analysis tool. This tool identifies statistically significant clusters with high or low values and outliers that are statistically different than their neighbors both in space and time[8] .

For the analysis to be statistically significant, the tool needs to perform permutations to determine the likelihood of the observed spatial distribution resulting from chance. Due to the size of the dataset and the minimum number of permutation required (99), the analysis was performed on a cube of 500 meters and 12 hour intervals.

Figure 19 shows the percentage of statistically significant clusters of high values of $PM_{2.5}$ over time in the study area (HH: high value of $PM_{2.5}$ surrounded by other high values). Similarly to the Hot Spot analysis, the area around El Bosque station is the one with highest percentages of high value clusters (associated to high $PM_{2.5}$ concentrations).

For bins of 12 hours, the maximum concentrations of $PM_{2.5}$ vary between 92 and 342 $\mu g/m^3$, with an average of the maximum values for the whole year of 158 $\mu g/m^3$.

[8] http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/localoutlieranalysis.htm

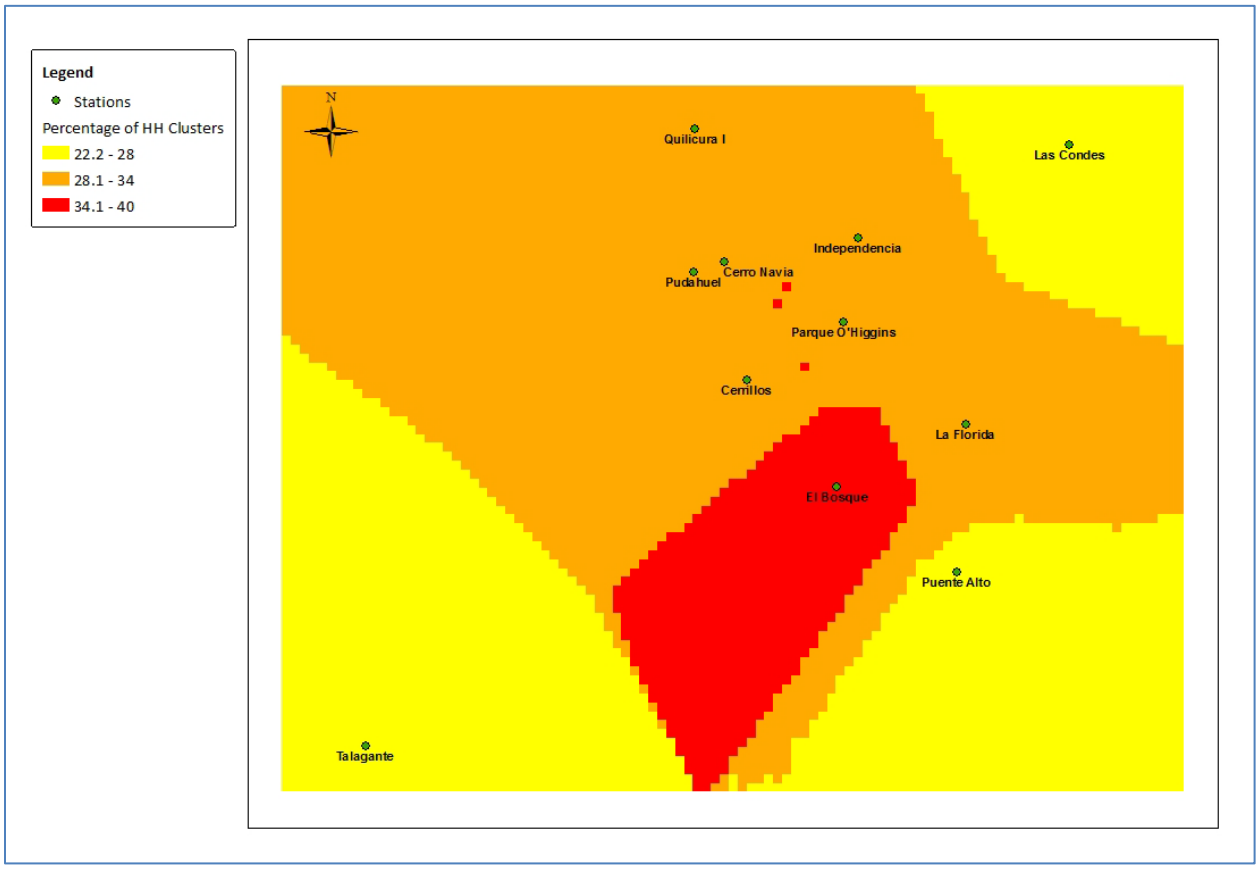


Figure 19: Percent of statistically significant clusters of high values of PM $_{2.5}$ concentrations

Figure 20 shows the number of time steps a bin was classified as a high cluster. The maximum (as the analysis was performed on a 12 hour cube) is 732. In the map, yellow indicates bins that were high clusters between 81 and 90 days in the year, orange between 91 and 120 days and red between 121 and 145 days. The whole study area is exposed to high concentrations of $PM_{2.5}$ for at least 81 days in a year (162 bins of 12 hours each). Areas surrounding Cerro Navia, El Bosque and La Florida stations are the ones that were exposed to high concentrations of $PM_{2.5}$ for the highest number of days during the year.

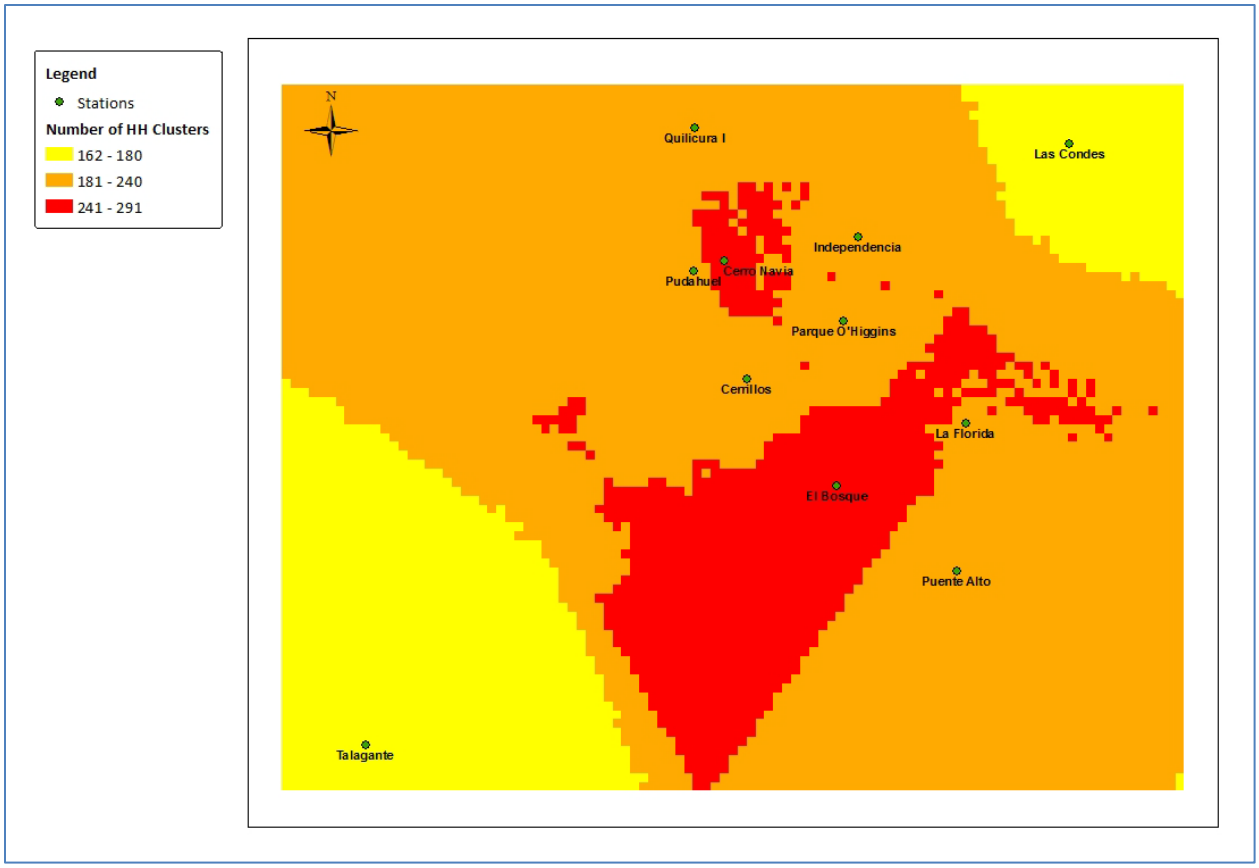


Figure 20: Number of statistically significant cluster of high values of $PM_{2.5}$ concentrations.

The Local Outlier Analysis tool adds its results to the space time cube, so this variability through time can be seen when the space time cube is visualized for local outliers, as shown in figure 21. In this case, the visualization of the cube was done using cube of 500 x 500 meters and 24 hour time intervals.

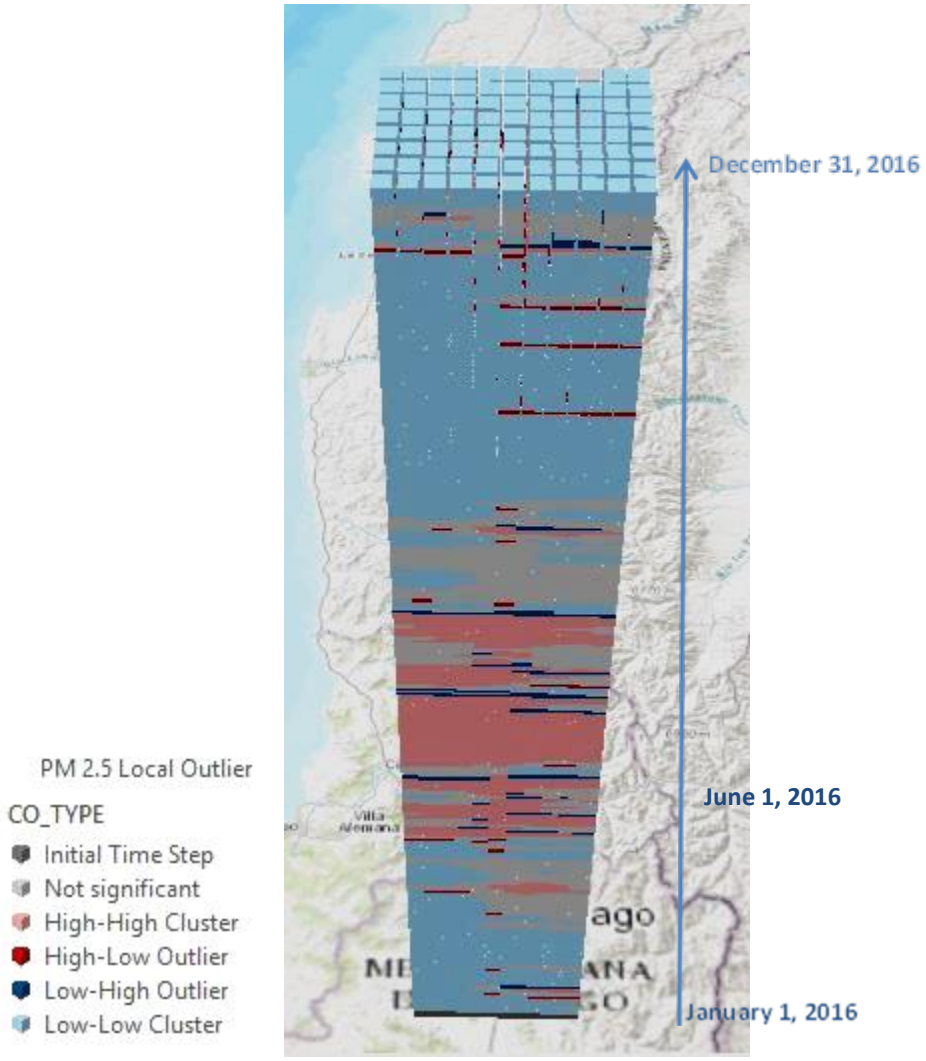


Figure 21: Local Outlier in the space time cube for $PM_{2.5}$ in in 3D.

Figure 22 shows the following maps:

- Top left: annual mean of $PM_{2.5}$ concentrations.
- Top right: median household income for 2009 and population below poverty level in 2015 (in red numbers).
- Bottom left: number of bins classified as part of high-high clusters (high values surrounded by high value neighbors), measured in days.
- Bottom right: population density in 2016 (people per hectare).

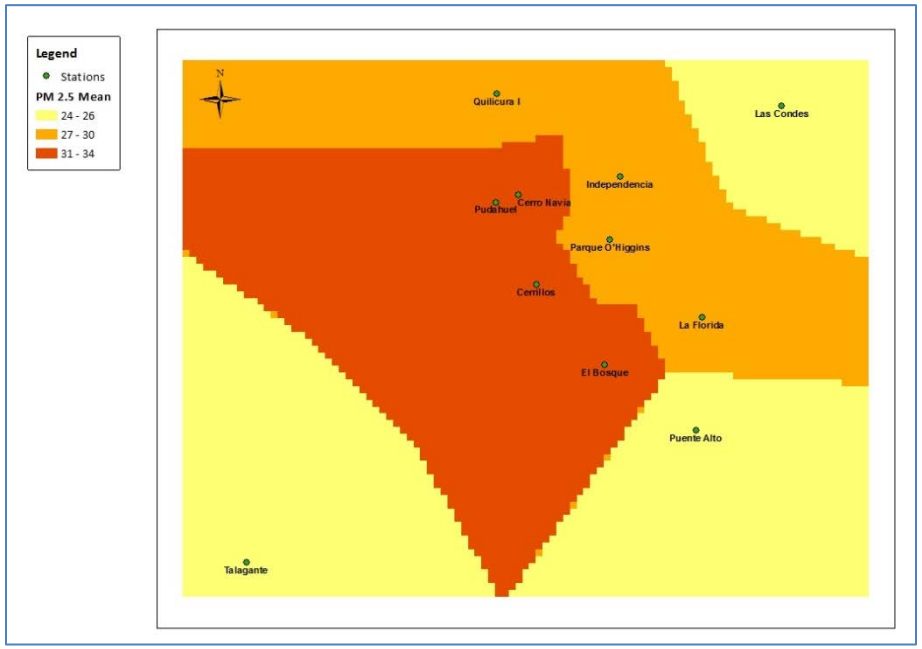
Figure 22 shows that areas with high average annual $PM_{2.5}$ concentrations show a general correspondence with areas of the city that have high population density and low median household

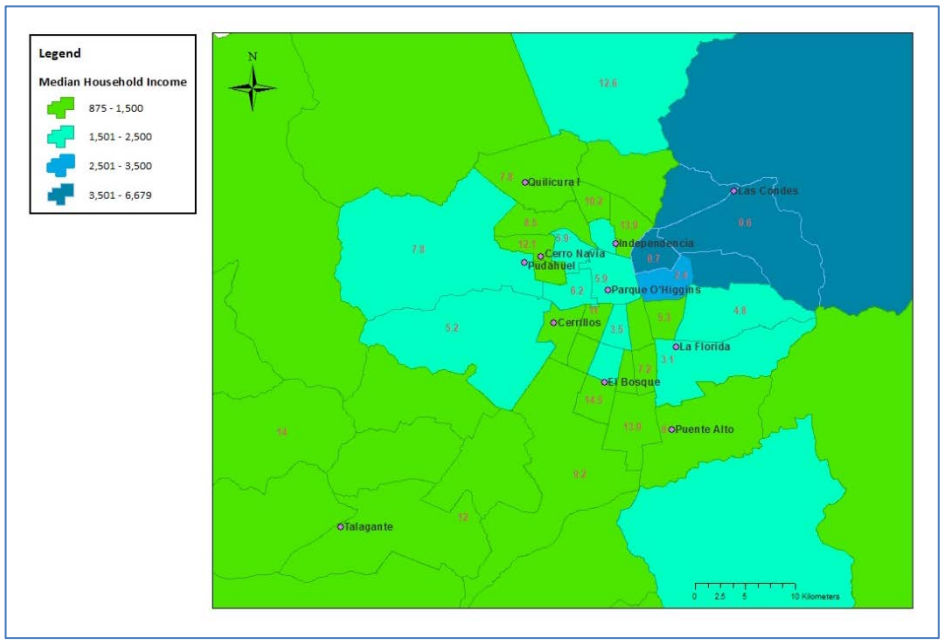
incomes. The median household income map also shows the percentage of population classified as poor by income (shown in red numbers) for the municipalities where the data was collected (CASEN 2015[9]).

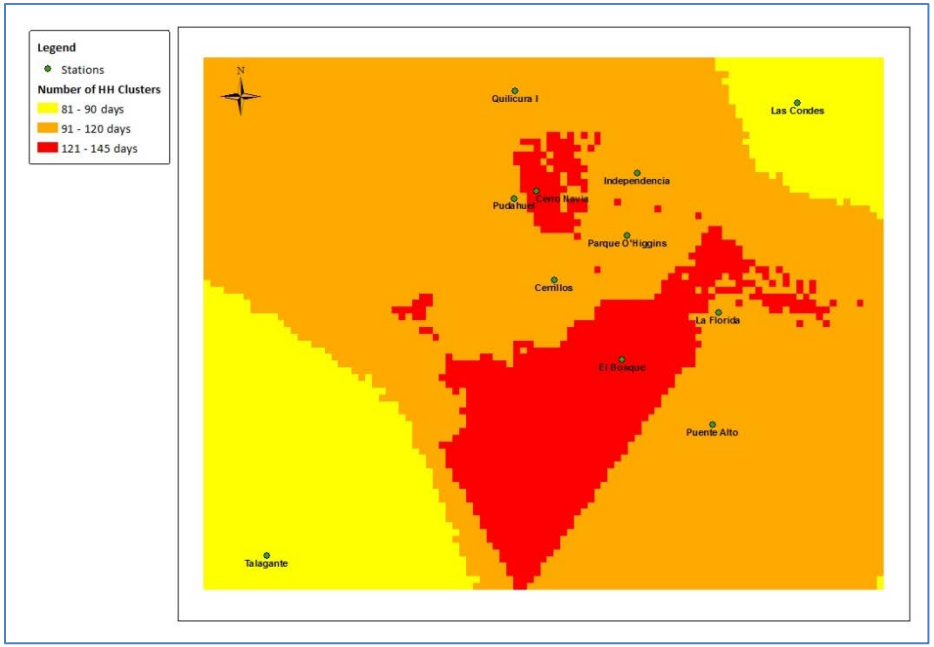
This correspondence indicates that a large part of Santiago's population is exposed to high levels of PM2.5 and that this portion of the population likely relies on public health care, which usually unable to cope with high demand in winter. This can lead to delays in receiving medical attention and worsening on chronic respiratory conditions. The results of this study can help to prioritize allocation of health care resources, improvement of city infrastructure and indicate that more aggressive measures are needed to achieve Chilean and WHO guidelines.

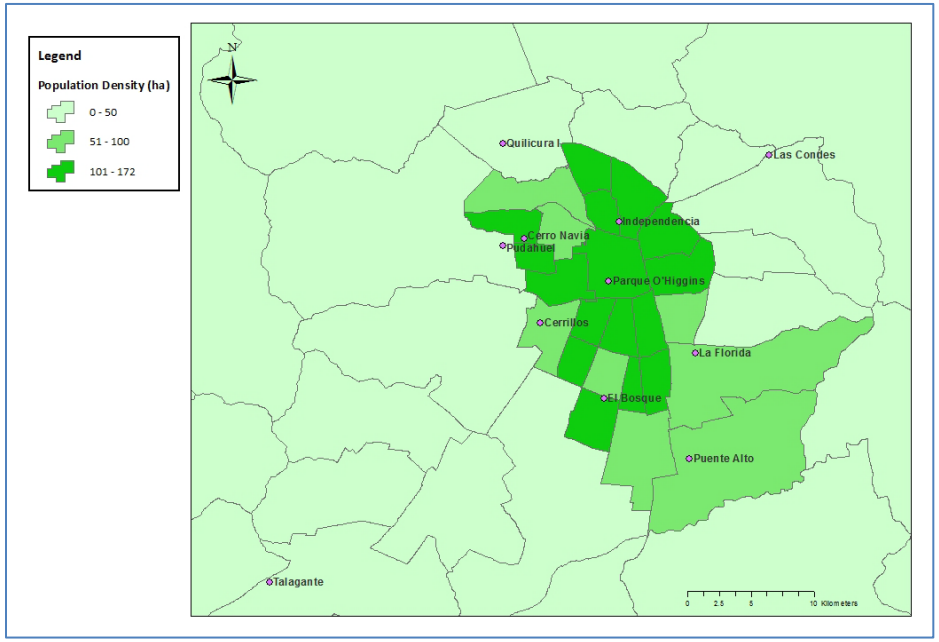
[9]Page 16 https://www.gobiernosantiago.cl/wp-content/uploads/2014/12/DOCUMENTO-POBREZA-Y-DISTR-ING-RMS-CASEN-2015.pdf.

Figure 22: Population density and median household income in US dollars.









CONCLUSIONS

$PM_{2.5}$ concentrations varies through the day, week and seasonally. Concentrations are especially high in winter months and most variable in spring.

Variographic analysis shows that $PM_{2.5}$ concentrations are highly correlated in time and space. General relative semivariograms clearly show that $PM_{2.5}$ concentration had a main 24 hour cycle and a secondary cycle of 10 to 12 hours for several stations, probably related to traffic patterns.

Environmental variables have negative coefficients but very weak predictive or explanatory power. Highest coefficients were for temperature and wind speed.

Kriging surfaces show areas where $PM_{2.5}$ concentrations are highest, which tend to be in the western part of the city. These areas have higher population density and lower income levels.

While the Emerging Hot Spot analysis found zones in the $PM_{2.5}$ concentrations and identified oscillating hot and cold spots. The Local Outliers Analysis determined that the study area was exposed to high levels of $PM_{2.5}$ for at least 80 days during 2016.

REFERENCES

- Cifuentes, L., Krupnick, A., O'Ryan, R., & Toman, M. 2005. Urban Air Quality and Human Health in Latin America and the Caribbean. Inter-American Development Bank. Retrieved October 31, 2016 from https://publications.iadb.org/handle/11319/2988 .
- Deutsch, C. & Journel, A. 1998. GSLIB Geostatistical Software Library and User´s Guide. New York: Oxford University Press.
- Deutsch, J. 2015. Geostatistics Lessons. Retrieved October 31, 2016 from http://www.geostatisticslessons.com/
- ESRI. 2016. Tool help documentation for ArcGIS Desktop 10.4.
- Gramsch, E., Reyes, F., Oyola, P., Rubio, M.A., López, G., Pérez, P. and Martínez, R. 2014. Particle size distribution and its relationship to black carbon in two urban and one rural site in Santiago de Chile. Journal of the Air & Waste Management Association 64: 785-796.
- Gramsch, E., 2014. Detailed Inventory of Santiago Air Pollution press release. South Pacific Review. Retrieved October 26, 2016 from http://southernpacificreview.com/2014/07/25/university-of-santiago-publishes-detailed-inventory-of-santiago-air-pollution/
- Green, J. and Sanchez, S. 2013. Air Quality in Latin America: An Overview. Clean Air Institute. Retrieved October 21, 2016 from http://www.cleanairinstitute.org/calidaddelaireamericalatina/cai-report-english.pdf
- Jhun I., Oyola, P., Moreno F., Castillo, M. and Koutrakis P. 2013. $PM_{2.5}$ mass and species trends in Santiago, Chile 1998 to 2010: The impact of fuel-related interventions and fuel sales. Journal of the Air & Waste Management Association 63:161-169.
- Li, H., Guo, B., Han, M., Tian, M. and Zhang, J. 2015. Particulate Matters Pollution Characteristic and Correlation between PM ($PM_{2.5}$, PM_{10}) and Meteorological Factors during the Summer in Shijiazhuang. Journal of Environmental Protection 6:457-463.
- Ministerio del Medio Ambiente (MMA). 2014. Estudio "Actualización y sistematización del inventario de emisiones de contaminantes atmosféricos en la Región Metropolitana. Retrieved October 31, 2016 from http://www.sinia.cl/1292/articles-56914_Inf_Inventarios_FINAL.pdf
- Ministerio del Medio Ambiente (MMA). 2011. Reporte 2005 -2009. Registro de Emisiones Contaminantes, RECT. Departamento de Estadísticas e Información Ambiental, División de Estudios del Ministerio del Medio Ambiente. Retrieved June 28, 2016 from http://www.retc.cl/wp-content/uploads/2015/07/articles-56962_Quinto_reporte_RETC.pdf

- Ministerio del Medio Ambiente (MMA). 2015. Segundo Reporte del estado del Medio Ambiente, 2015. Retrieved October 15, 2016 from http://sinia.mma.gob.cl/

- Pitard, F. 1993. Pierre Gy'S Sampling Theory and Sampling Practice. CRC Press.

- Perez, P. & Gramsch, E. 2016. Forecasting hourly $PM_{2.5}$ in Santiago de Chile with emphasis on night episodes. Atmospheric Environment 124: 22-27.

- Power, M., Kioumourtzoglou, M., Hart, J., Okereke, O., Laden, F., and Weisskopf, M. 2015. The relation between past exposure to fine particulate air pollution and prevalent anxiety: observational cohort study. British Medical Journal (BMJ) 2015;350:h1111. Retrieved November 7, 2016 from http://www.bmj.com.ezaccess.libraries.psu.edu/content/bmj/350/bmj.h1111.full.pdf

- Riojas-Rodríguez H., Soares da Silva A., Texcalac-Sangrador JL. And Moreno-Banda GL. 2016 Air pollution management and control in Latin America and the Caribbean: implications for climate change. Rev Panam Salud Pública. 2016;40(3):150–59. Retrieved November 30, 2016 from http://iris.paho.org/xmlui/handle/123456789/31229

- Shah, A., Lee, K., McAllister, D., Hunter, A., Nair, H., Whiteley, W., Langrish, J., Newby, D. & Mills, N. 2015. Short term exposure to air pollution and stroke: systematic review and meta-analysis. British Medical Journal (BMJ) 2015;350:h1295. Retrieved November 7, 2016 from http://www.bmj.com.ezaccess.libraries.psu.edu/content/bmj/350/bmj.h1295.full.pdf

- Sherman, C. 2015 Air Pollution Also Not Good For Your Brain. The Dana Foundation. Retrieved November 1, 2016 from http://www.dana.org/News/Air_Pollution_Also_Not_Good_For_Your_Brain/

- World Health Organization regional office and European Centre for Environment and Health Bonn Office. 2006. Health risks of particulate matter from long-range transboundary air pollution. Retrieved October 21, 2016 from http://www.euro.who.int/__data/assets/pdf_file/0006/78657/E88189.pdf

- World Health Organization regional office for Europe. 2013. Health effects of particulate matter: Policy implications for countries in Eastern Europe, Caucasus and central Asia. Retrieved October 21, 2016 from http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf

- Zirui, L., Hu B., Wang L., Wu F., Gao W., and Wang, Y. 2015. Seasonal and diurnal variation in particulate matter (PM_{10} and $PM_{2.5}$) at an urban site of Beijing: analyses from a 9 year study. Environmental Science and Pollution Research 22:627-642.

DATA SOURCES:

- National Information System on Air Quality: http://sinca.mma.gob.cl/index.php/region/index/id/M
- USGS earth explorer: http://earthexplorer.usgs.gov