Uncovering Income Segregation and its Relationship to Pollution ($PM_{2.5}$) in Vancouver, British Columbia

Geography 418: Advanced Spatial Analysis

Instructor: Dr. Christopher Bone

TA: Jason Kelley

By: Cole Glover – V00838832

Submit by: 23:55 on Dec. 6, 2019

---

1 Introduction

In recent decades, numerous studies suggest a temporal relationship between income and pollution to exhibit the inverted U-shape behavior of an environmental Kuznets curve – where pollution first increases, but later falls with increasing income (Carson *et al.*, 1997, McKitrick and Wood, 2017). Whereas most of the aforementioned studies are performed on provincial- and state-wide scales, fewer studies focus on a city-scale. Vancouver, British Columbia, as with many other large metropolitan areas in Canada, is experiencing increased spatial polarization of higher- and lower-income neighbourhoods (Breau *et al.*, 2018) and is responsible for high concentrations of primary air pollutants (Setton, 2007). By employing various statistical analyses on cross-sectional data, this study aims to identify if the spatial distribution and concentration of $PM_{2.5}$ can be adequately predicted by median income within the Greater Vancouver Regional District.

2 Methods

2.1 Study Area

The analyses performed in this study encompasses the area within the Greater Vancouver Regional District (GVRD; Figure 1). Within this metropolitan area, neighbourhoods have been defined by dissemination area (DA) boundaries, which consists of 3363 DAs, gathered from Census Canada in 2016**.** Income data from 2016, sourced by Census Canada, is also collected at the DA level. Additionally, the GVRD's industrial sectors, airport, shipping ports, and large traffic volumes makes ambient fine particulate matter ($PM_{2.5}$) a primary cause for concern. $PM_{2.5}$ (ppm) data used in this study, provided by the Canadian Urban Environmental Health Research

Consortium (CANUE), is annually averaged from 2016 and collected at the postal code level from 160 sample sites (Figure 2). As multiple pollution sample sites existed within a DA, the maximum $PM_{2.5}$ value of each DA is chosen to represent the DA the sample site exists in.

2.2 Exploring income and pollution distribution

2.2.1 Spatial Autocorrelation - Moran's *I* Analysis

Firstly, to identify the distribution of high and low median income in the GVRD, a Local Moran's *I* is computed. By utilizing a "Queens Weight" matrix, the Local Moran's *I* determines a geographic relationship between dissemination areas and then calculates the relationship between each observation and the weighted averaged between its eight nearest neighbours (Breau *et al.*, 2018). The Local Moran's *I* analysis uncovers regions of spatial autocorrelation within the study site.

2.2.2 Spatial Interpolation – Universal Kriging

To create a predicted $PM_{2.5}$ (ppm) surface, universal kriging (UK) is employed by using 160 random pollution observations from across the GVRD. The UK model integrates a simple second-order polynomial trend surface (accounting for basic trends which likely underlie the pollution data) with a spherical semivariogram model fitted through residuals. In comparison to model-fits, the spherical model most adequately represents spatial variance of pollution residuals; however, various models exist which describe different spatial variance trends. Spatial variance and 95% confidence interval maps were produced to evaluate accuracy of the interpolated surface. Although UK utilizes a weighted sum of local values, UK presents itself as the optimal interpolation method, as it avoids biases accompanied by a user-assigned distance-weighting scheme, as seen in inverse distance weighting.

2.3 Determining Relationships – Regression Analysis

Here, an ordinary least squares (OLS) regression model is used to regress $PM_{2.5}$ on median income in the GVRD. In the OLS, the dependent variable ($PM_{2.5}$) is assumed to be a linear function of the independent variable (median income). To minimize the sum of error square of unbiased observations, the model should be able to predict expected values of the dependent variable (Shrestha, 2006); furthermore, unbiased observations are based on the following assumptions: normality, homogeneity and independence of residuals (Montgomery *et al*., 2001). Violations of these assumptions produces an inaccurate and biased estimator (Shrestha, 2006). To postulate that data used in this study does not violate any aforementioned assumptions would be unrealistic. As such, further methods are used to investigate violation occurrence and a potential work-around, should a violation occur. Please note: before running the OLS regression model, "NA" pollution values were removed from the pollution dataset.

2.3.1 Global Moran's *I* Analysis on Residuals

In order to determine if the study data violates any aforementioned assumptions of the OLS model, namely, the spatial independence of residuals, a Global Moran's *I* analysis is performed on residual values of OLS model. By performing a Global Moran's *I* on residuals, it can be determined if residual values are clustered and exhibit dependence on each other, thus violating OLS assumptions.

2.3.2 Geographically Weighted Regression

As spatial data are often not stationary and relationships between variables may be localized or vary across space (Hanham and Spiker, 2005), a geographically weighted regression (GWR) evaluates multiple regressions between the independent and dependent variables at each sample site, capturing varying spatial relationships (Donkelaar *et al.*, 2015). Furthermore, when fitting individual regressions at sample points across the study site, the GWR applies a proximity-weighting function (known as a kernel bandwidth) to neighbouring observations – this assigns greater influencing weight to proximal sample points (Shrestha, 2006). However, through cross-validation, this manipulable bandwidth was optimized by minimizing the bandwidth's root mean square error (RMSE).

2.4 Point Pattern Analysis

Lastly, the kernel density estimator (KDE) is employed to produce a point-density surface. The KDE is a power tool to visualize spatial patterns of point data to detect sample site hotspots. Since the accuracy of KDE hinges on its search radius, the search radius of the KDE is also optimized by minimizing the RMSE using cross-validation.

Please find the R script used to perform aforementioned methods at: https://github.com/tropicole

3 Results

3.1 Distribution of Median Income & $PM_{2.5}$

To display the distribution of income across the GVRD a simple choropleth map of median income is created (Figure 3). Regions of high median income are seen in North and West Vancouver, Kitsilano, Tsawwassen, and Point Grey. Low income regions are seen in Richmond, Surrey, and in surrounding areas of Burnaby.

3.1.1 Local Moran's *I*

Local Moran's *I* outputs display areas of nearby similarity and dissimilarity (Figure 4.1). Nearby regions of similarity are represented by large positive values, whereas, nearby regions of dissimilarity are represented by large negative values. Dissimilarity of median income across the GVRD is prominent. This is also represented in a Local Moran's *I* plot (Figure 4.2). Pockets of median income similarity are seen in West, North, and Downtown Vancouver, Kitsilano, Point Roberts, Cloverdale, Richmond, and South of Surrey.

3.1.2 Universal Kriging

Figure 5.1 presents a continuous surface of estimated $PM_{2.5}$ concentrations using the universal kriging model. The predicted pollution surface displays concentrations of $PM_{2.5}$ between 5-7 ppm on the east and west portions of the study site, from Maple Ridge to Langley and along the coast, from West Vancouver to Richmond. Low $PM_{2.5}$ concentrations of 0-1 ppm are seen throughout Burnaby toward Surrey. Variance and confidence interval surfaces are produced (Figure 5.2 and Figure 5.3, respectively) to suppl. Overall, the variance surface presents low values throughout the GVRD, though higher values exist on the northeast perimeter of the study site. Similarly, the 95% confidence interval map displays high measurement error associated with high variance regions. Areas of low measurement error are typically found near sample sites.

3.3 Regression Analysis & Global Moran's *I*

Residual outputs of the OLS do not appear to be strongly symmetrical (i.e. normally distributed), nor do error outputs appear to be relatively low (residual standard-error of 1.299 and an adjusted R-squared value of 0.0285). A relatively poor OLS performance, combined with a strikingly high Moran's *I* statistic (0.9737032347, range: [-1.000000,1.044212]) and Z-score (88.13916) indicates clear signs of spatial autocorrelation in the residuals of the $PM_{2.5}$ study data.

3.4 GWR model

In comparison to a set of values which represent and entire study area, as seen in the OLS, the GWR model computes a set estimated parameters at each sample point (Shrestha, 2006). Outputs include an AIC of -5600.026, indicating an optimized bandwidth used in the GWR. In addition, GWR coefficient and R-squared values are mapped across the study area (Figure 6.1 and Figure 6.2, respectively). The resultant GWR coefficient surface displays the direction and magnitude of change in the dependent variable ($PM_{2.5}$) in response to change in the independent variable (median income). In regards to significance of GWR coefficient values, the local R-square plot displays strong significance across the study area.

3.5 Point Pattern Analysis

The resultant heat map (Figure 7) displays clustering of sample sites throughout the central Vancouver and Burnaby region. The cross-validation tool, used to optimize search bandwidth, determined a search radius of 2 km.

4 Discussion

Descriptive statistics were performed to describe the relationship between pollution and income within the GVRD. As seen by Figure 3 and Figure 4, distinction in Vancouver neighbourhood income appears to be sharp and dissimilar across the region, similar to findings by Breau *et al.* (2018). The distribution of pollution throughout the GVRD, as estimated by the UK model, increases outwards from the central region of the study area (Figure 5.1). Significantly high regions of $PM_{2.5}$ concentrations (Figure 5.3) are prominent throughout central Vancouver, Kitsilano, and Richmond. Due to the nearby presence industrial sectors, high-concentration roadways, Vancouver's main airport, and shipping ports (Burrard Inlet), as well as population-dense areas (central Vancouver), these regions are known for high concentrations of air pollution (Setton, 2007). Similar to a majority of literature, this study uses median income as an independent variable to predict $PM_{2.5}$ concentrations (Akbostancı *et al.*, 2009, Carson *et al.*, 1997, McKitrick *et al.*¸2017). As such, $PM_{2.5}$ was regressed on median income in the GVRD; however, by performing a Global Moran's *I* test on $PM_{2.5}$ residuals, it was found that the residuals exhibit prominent clustering (Z-score = 88.13916), thus a spatial non-stationarity of the relationship in question, proving that the data violate assumptions of independence and ultimately invalidate the OLS model. As such, a GWR model is used to describe the localized relationships between median income and $PM_{2.5}$ concentrations.

Shown by Figure 6.2, the GWR model performed well, produced high R-squared values (0.810-0.995) throughout nearly all of the GVRD. When coincided with regions of high R-squared values, GWR coefficients can predict or at least aid in understanding the factors that contribute to dependent variable outcomes. When considering the GWR coefficient surface, it is clear that the relationship between income and $PM_{2.5}$ varies across the study area (Figure 6.1). In regions such as Surrey and Richmond, when median income is considerably high (Figure 3) and $PM_{2.5}$ is relatively low (Figure 5.1), the GWR coefficients indicate that per unit increases in median income result in a general decrease in $PM_{2.5}$ concentrations (Figure 6.1). In contrast, central Vancouver exhibits the opposite relationship – per unit increases in median income result in increases in $PM_{2.5}$ concentrations. Moreover, this variation in the relationship between income and $PM_{2.5}$ concentrations is prominent throughout the GVRD, indicating that finding a casual relationship in a cross-sectional study is not simple; furthermore, there are many factors which can play a role in pollution output and distribution, such as population density, location, density of industrial areas, and environmental effects. It is clear that the level of $PM_{2.5}$ cannot be determined directly by level of income; however, similar to work by Akbostancı *et al.*, 2009, Carson *et al.*, 1997, McKitrick *et al.*¸2017, this study presents median income an an important piece in determining pollution concentrations. To evaluate whether or not the relationship between median income and pollution throughout the GVRD exhibits the behaviour of an environmental Kuznets curve, further studies should encompass a multi-variant, temporal analysis of income and pollution within the GVRD.

Appendix

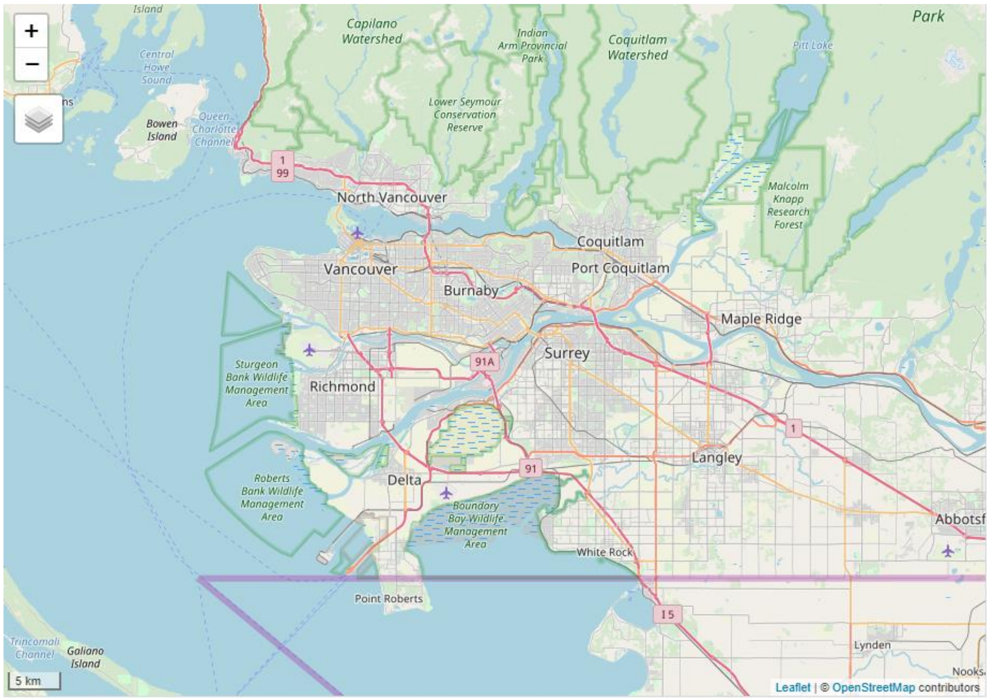

Figure 1. Reference map of the Greater Vancouver Regional District.

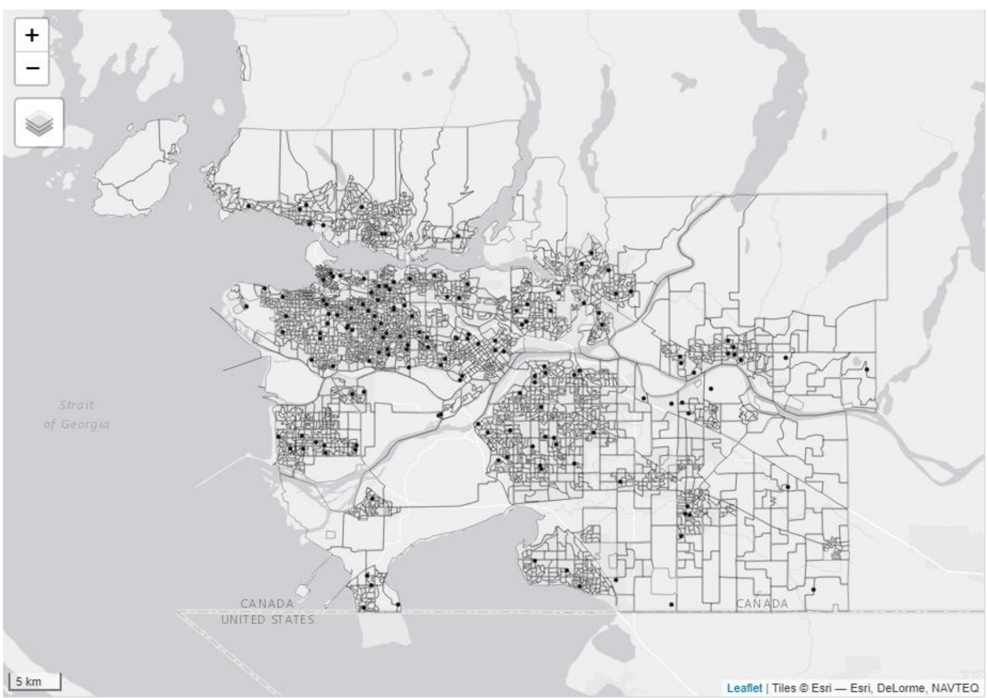

Figure 2. Study area boundary and location of randomly selected $PM_{2.5}$ samples sites.

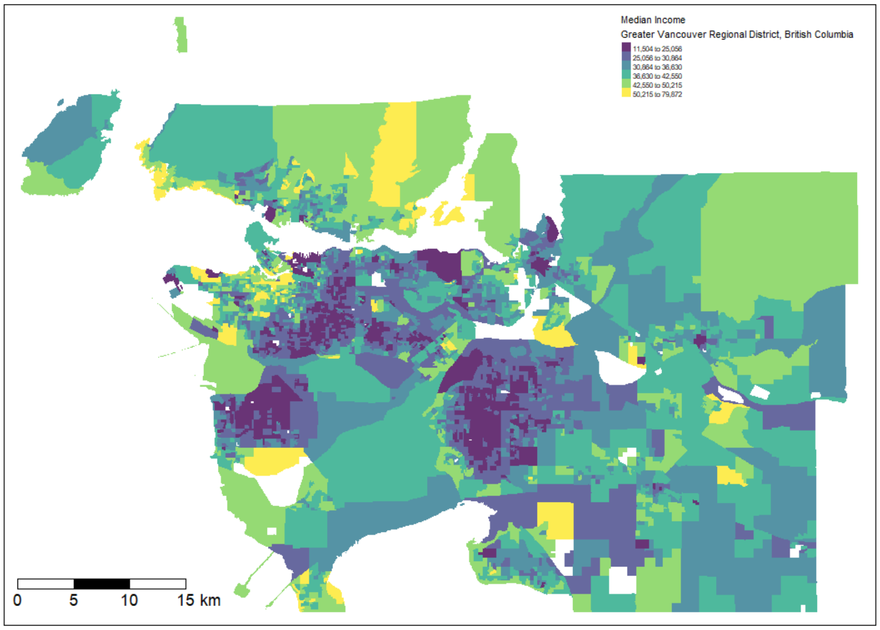

Figure 3. Median Income in 2016 in the Greater Vancouver Regional District.

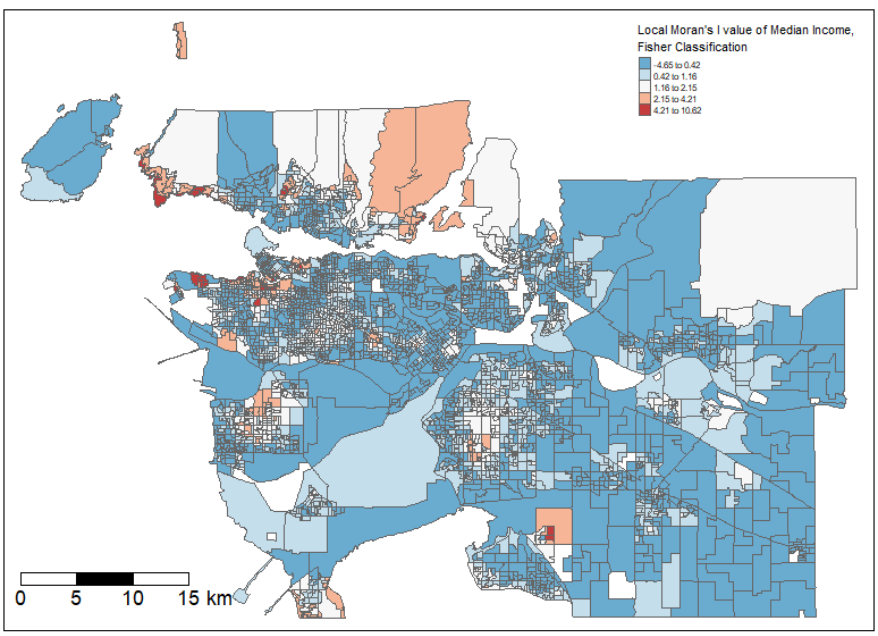

Figure 4.1. Local Moran's *I* analysis output for median income the GVRD. Shown by low *I* values (blue), median income is prominently dissimilar across the region. Similar median income values (red) reside in West, North, and central Vancouver, Point Grey, Richmond, and Cloverdale.

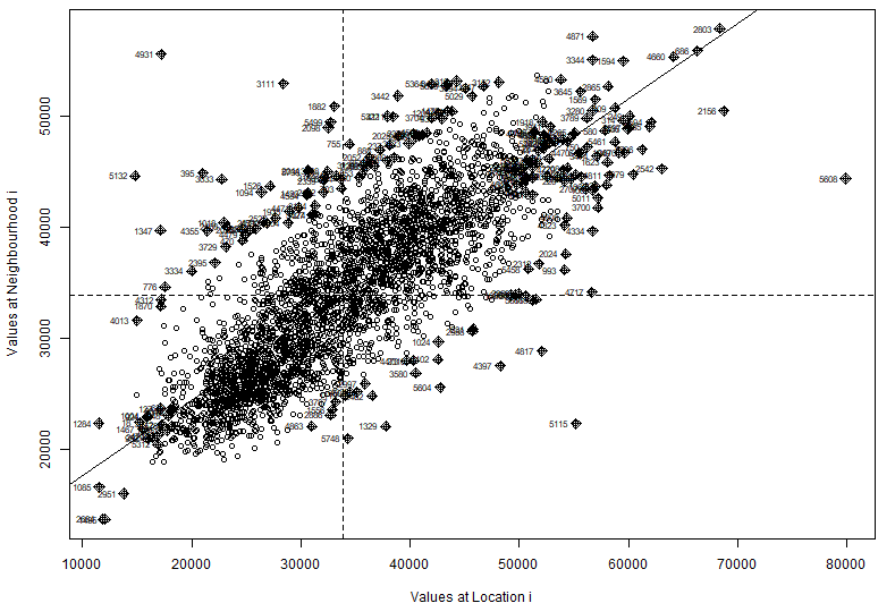

Figure 4.2. Local Moran's *I* scatterplot for median income in the GVRD. The plot displays prominence of dissimilar values, seen in quadrant III. The range of points in quadrant I represents the large range of median incomes in the GVRD. The positive behavior of the 'best-fit' line infers positive spatial autocorrelation without the study area.

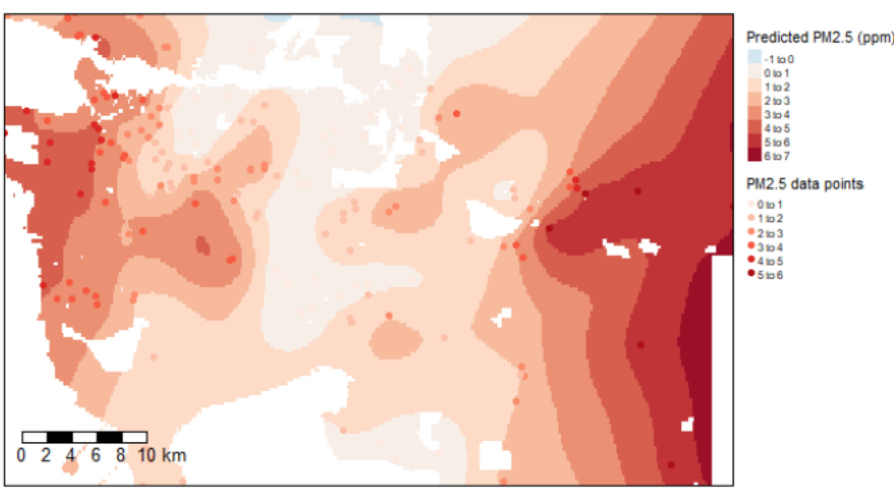

Figure 5.1 Estimated $PM_{2.5}$ (ppm) concentration surface using a universal kriging, governed by 160 randomly selected sample points, a second-order polynomial trend surface, and a spherical variogram model.

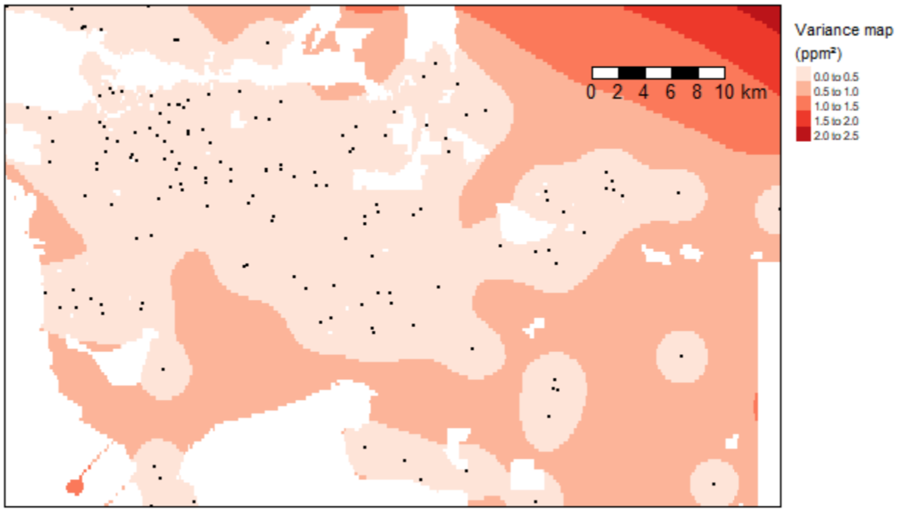

Figure 5.2 Universal kriging variance surface, governed by a second-order polynomial function and spherical variogram model. Smaller variances represent a more accurate estimated surface.

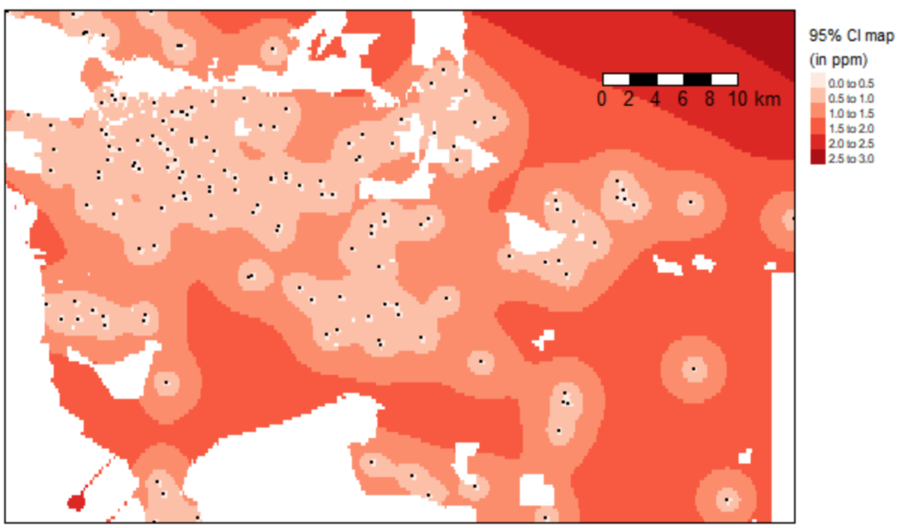

Figure 5.3 Universal kriging confidence interval map. Each pixel represents a range of $PM_{2.5}$ concentrations around an estimated point at 95% confidence. Darker reds indicate a wider range of potential values around an estimated point, whereas lighter colours indicate a narrower range of values around an estimated point.

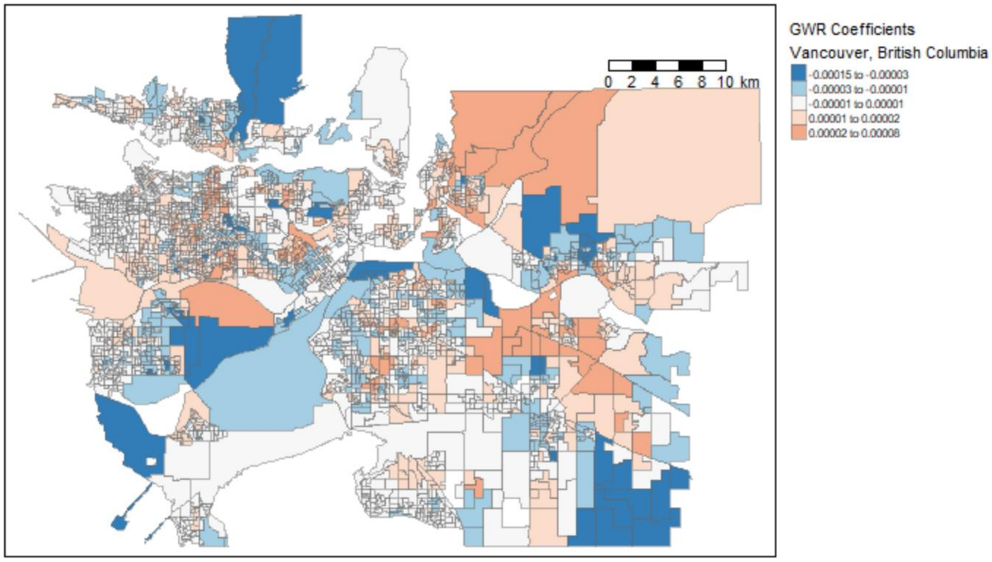

Figure 6.1 Distribution of GWR coefficients throughout the GVRD, representing magnitude and direction of change in the dependent variable, in response to change of the independent variable. The opacity of colour is proportional to the magnitude of change in response to perturbations in the independent variable.

Figure 6.2 Distribution of local R-square values across the GVRD. Darker greys indicate a lower R-square value; however, a strong majority of the study area exhibits a high R-square value, as indicated by lighter shade of grey.

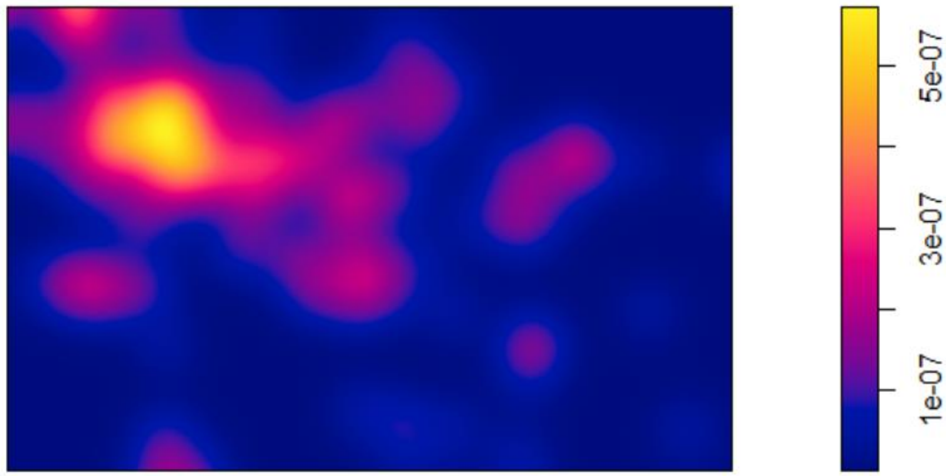

Figure 7. Kernel Density Estimation heat map of sample sites through the GVRD. An optimized search bandwidth of 2 km was used. Prominent cluster of sample sites is seen throughout central Vancouver and Burnaby.

References

Akbostancı, E., Türüt-Aşık, S., & Tunç, G. I. (2009). The relationship between income and environment in Turkey: Is there an environmental Kuznets curve? Energy Policy, 37(3), 861–867. doi: 10.1016/j.enpol.2008.09.088

Breau, S., Shin, M., and Burkhart, N. (2018). Pulling Apart: New Perspectives on the Spatial Dimensions of Neighbourhood Income Disparities in Canadian Cities. *Journal of Geographical Systems*, 20(1), 1-25. doi:1007/s10109-017-0255-0

Carson, R.T., Jeon, Y., and McCubbin, D. (1997). The Relationship Between Air Pollution Emissions and Income: US Data. *Environment and Development*, 2(4), 433-450. doi:10.1017/s1355770x97000235

Census Canada (2019). Retrieved November 25, 2019, from https://www150.statcan.gc.ca/n1/en/type/data?MM=1.

Donkelaar, A.V., Randall, M.V., and Robert, J.D. (2015). High-Resolution Satellite-Derived PM2.5 from Optimal Estimation and Geographically Weighted Regression over North America. *Environmental Science & Technology*, 49(17), 10482-10491. doi: 10.1021/acs.est.5b02076

Hanham, Robert & Spiker, J. (2005). Urban Sprawl Detection Using Satellite Imagery and Geographically Weighted Regression. *Geo-Spatial Technologies in Urban Environments*. doi:10.1007/3-540-26676-3_12.

McKitrick, R. and Wood, J. (2017). An Examination of the Relationship between Air Quality and Income in Canada. *Canadian Journal of Agriculture Economics*, 65, 69-92. doi: 10.1111/cjag.12101

Montgomery, D.C., Peck, E.A., Vining, G.G. (2001). Introduction to Linear Regression Analysis, *3rd edition*. *John Wiley & Sons, Inc*.

Shrestha, P.M. (2006). Comparison of Ordinary Least Square Regression, Spatial Autoregression, and Geographically Weighted Regression for Modeling Forest Structural Attributes Using a Geographical Information System (GIS)/Remote Sensing (RS) Approach. *Master thesis - University of Calgary*, 37.

Setton, E.M. (2007). Probabilistic Estimates of Variability in Exposure to Traffic-related Air Pollution in the Greater Vancouver Regional District – A Spatial Perspective. *Master thesis - University of Victoria, 3.*

Setton, E.M. (2019, October 18). The Canadian Urban Environmental Health Research Consortium (CANUE). Retrieved from https://canue.ca/.