Improving the Accuracy of Urban Environmental Quality Assessment Using Geographically-Weighted Regression Techniques

by

Kamil Fasial

B.Sc., King Abdul Aziz University, Jeddah, Saudi Arabia, 2006

M.A.Sc., Ryerson University, Ontario, Canada, 2011

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Civil Engineering

Toronto, Ontario, Canada, 2017

©Kamil Fasial, 2017

Author's Declaration

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Improving the Accuracy of Urban Environmental Quality Assessment Using Geographically-Weighted Regression Techniques

Doctor of Philosophy 2017

Kamil Fasial

Civil Engineering

Ryerson University

Abstract

Population growth around the world may cause an adverse impact on the environment and the human life. Thus, modeling the Urban Environmental Quality (UEQ) becomes indispensable for a better city planning and an efficient urban sprawl control. To evaluate the impact of city development, this study aims to utilize remote sensing and Geographic Information System (GIS) techniques to assess the UEQ in two major cities in Ontario, Canada.

The main objectives of this research are: 1) to examine the relationship of multiple UEQ parameters derived from remote sensing, GIS and socio-economic data; 2) to evaluate some of the existing methods (e.g. linear regression, GIS overlay and Principal Component Analysis (PCA)) for assessing and integrating multiple UEQ parameters; 3) to propose a new method to weight urban and environmental parameters obtained from different data sources; 4) to develop a new method to validate the UEQ results with respect to three socio-economic indicators.

Remote sensing, GIS and census data were first obtained to calculate various environmental, urban parameters and socio-economic indicators. The derived parameters and indicators were tested to emphasize their relationship to UEQ. Three Geographically-Weighted Regression (GWR) techniques were used to integrate all these environmental, urban parameters and socio-economic indicators. Three key indicators including family income, the level of education and land value were used as a reference to validate the outcomes derived from the integration techniques. The results were evaluated by assessing the relationship between the extracted UEQ results and the three indicators. The findings showed that the GWR with spatial lag model represents an improved precision and accuracy up to 20% with respect to GIS overlay and PCA techniques. The final outcomes of the research can help the authorities and decision makers to understand the empirical relationships among regional science, urban morphology, real estate economics and economic geography.

Acknowledgements

Above all, I am grateful to Allah who has given me the strength, health, patience, and success to complete my graduate studies at Ryerson University.

I owe tremendous thanks to many people whose assistance was indispensable in completing this research endeavor. First among these is Dr. Ahmed Shaker, my supervisor. Dr. Shaker is an extremely helpful and down to earth person, who gave me encouragement, valuable advice not only on my studies but also regarding everything that he found which was important to my family and me. Dr. Shaker was always there for me no matter what or when even with his extremely busy schedule. Also, special thanks and appreciation go to Dr. Wai Yeung Yan, a Postdoctoral fellow at Ryerson University for full support, encouragement and valuable ideas. My dissertation wouldn't have been achievable without the knowledge and effort that I received from him regarding geographic information systems and remote sensing.

In addition, I wish to extend my gratefulness and thanks to members of my Ph.D. examining committee, Prof. Mir Abolfazl Mostafavi, Prof. Songnian Li, Dr. Raktim Mitra, and Dr. Darko Joksimovic, for their valuable time and effort to review my dissertation.

Special thanks also are extended to Mr. Dan Jakubek and Noel Damba from Geospatial Map and Data Centre in the Ryerson Library who provided technical guidance, support, and provision of all the GIS data. I also acknowledge the United States Geological Survey (USGS) for the remote sensing data and information provided through the EarthExplorer platform.

Special thanks also are given to King Abdul-Aziz University and the Saudi Culture Bureau. They both gave me the opportunity to study abroad at Ryerson University with full financial support and assistance. Moreover, I am also grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and the Association of Ontario Land Surveyors (AOLS) for their indirect financial support. Finally, I would like to thank the people who contribute the most to my life including my parents, my wife and my daughter, Toleen. Without them, none of this work would have been possible.

Table of Contents

Declaration **ii**

Abstract **iii**

Acknowledgements **v**

List of Tables **ix**

List of Figures **x**

1 Introduction **1**

1.1 Research Motivation 1

1.1.1 UEQ parameters 3

1.2 Research Problem 6

1.3 Research Objectives 7

1.4 Dissertation Workflow 8

1.5 Dissertation Outline 9

2 Modelling the Relationship between the Gross Domestic Product and Built-Up Area **13**

2.1 Abstract 13

2.2 Introduction 14

2.3 Datasets and Methods 18

2.3.1 Datasets 18

2.3.2 Methodology 19

2.4 Results and Discussion 24

2.4.1 Built-Up Areas 24
2.4.2 Regression Analysis between the Socio-Economic indicators and Built-Up Areas 27
2.4.3 Discussion 32
2.5 Conclusions 33

3 An Investigation of GIS Overlay and PCA Techniques for UEQ Assessment: A Case Study in Toronto, Ontario, Canada 35

3.1 Abstract 35
3.2 Introduction 36
3.3 Datasets 40
3.4 Methodology 42
3.4.1 Environmental Parameters 43
3.4.2 Urban Planning Parameters 47
3.4.3 Socio-Economic indicators 50
3.4.4 Ranking the Parameters 51
3.4.5 Data Integration of Multiple Environmental and Urban Parameters 53
3.5 Results and Analysis 56
3.5.1 GIS Overlay Analysis 56
3.5.2 Principal Component Analysis 57
3.5.3 UEQ Validation Results 64
3.6 Conclusions 67

4 Improving the Accuracy of UEQ Assessment Using Geographically-Weighted Regression Techniques 70

4.1 Abstract 70
4.2 Introduction 71
4.3 Datasets 75
4.4 Methodology 76
4.4.1 Ranking the Parameters 77
4.4.2 Data Integration of Multiple Environmental and Urban Parameters 80
4.5 Results and Discussion 87
4.5.1 GIS Overlay Analysis 87

4.5.2 Principal Component Analysis 88
4.5.3 Geographically-Weighted Regression 91
4.5.4 UEQ Results Validation 94
4.6 Conclusions 103

5 Conclusions and Future Work **105**
5.1 Conclusions 105
5.2 Limitations and Future Work 110

References **118**

List of Tables

2.1 The data sources for the seven major cities. 19

3.1 The data sources for City of Toronto. 41
3.2 The sum of the socio-economic indicators. 56
3.3 The correlation coefficient matrix among all of the parameters derived using the pixel-based method. 59
3.4 The parameters vs. the components in the pixel-based PCA. 60
3.5 The correlation coefficient matrix among all of the parameters for the object-based method. 62
3.6 The parameters vs. the components in the object-based PCA. 63

4.1 The data sources for City of Toronto and City Of Ottawa. 76
4.2 The sum of the socio-economic indicators. 86
4.3 The correlation coefficient matrix among all of the parameters derived from the PCA method in the City of Toronto. 88
4.4 The correlation coefficient matrix among all of the parameters derived from the PCA method in the City of Ottawa. 89

List of Figures

1.1 The overall dissertation workflow. 9

2.1 The overall workflow for modelling the relationship between the GDP and built-up area. 20

2.2 A pictogram to demonstrate how the built-up area is derived. 22

2.3 Landsat derived built-up image (left) and land use map (right) for the City of Montreal. 23

2.4 An illustration of the built-up pixels overlaid on the land use map. 23

2.5 Percentage of built-up areas within the land use. (**a**) 2005; (**b**) 2006; (**c**) 2007; (**d**) 2008; (**e**) 2009; (**f**) 2010. 25

2.6 Relationship between the percentage of built-up areas and the industrial areas from 2005 to 2010. 27

2.7 Relationship between % of built-up areas and the socio-economic indicators from 2005 to 2010. (**a**) % of built-up areas *vs.* GDP; (**b**) % of built-up areas *vs.* total employment; (**c**) % of built-up areas *vs.* population. 28

2.8 Relationship between % of built-up areas and the socio-economic indicators without Edmonton and Calgary from 2005 to 2010. (**a**) % of built-up areas *vs.* GDP; (**b**) % of built-up areas *vs.* total employment; (**c**) % of built-up areas *vs.* population. 29

2.9 Relationship between the real GDP and industrial areas from 2005 to 2010. (**a**) City of Ottawa; (**b**) City of Vancouver. 29

2.10 Relationship between the real GDP and industrial areas from 2005 to 2010. (**a**) City of Toronto; (**b**) City of Montreal; (**c**) City of Edmonton; (**d**) Québec City; (**e**) City of Calgary; (**f**) all cities. 30

2.11 Real GDP *vs.* industrial areas without Edmonton and Calgary. 31

3.1 City of Toronto (the study area). 41
3.2 The overall workflow for investigating of GIS overlay and PCA techniques. 42
3.3 (**a**) NDVI image derived from Landsat image (raster data); (**b**) NDVI map after transformation (vector data); (**c**) population layer at the census tract level; (**d**) population layer after transformation to sub-neighbour level. . . 52
3.4 The GIS polygons of the parameters. 53
3.5 (**a**) The LST layer in degrees Celsius before ranking the parameter; (**b**) the ranking of LST after the normalization. 53
3.6 The summed up ranks for all of the parameters. 54
3.7 The UEQ derived using the GIS overlay method. 57
3.8 The UEQ derived using the first component of the pixel-based PCA method. 60
3.9 The UEQ parameters versus PCA Component 1. 61
3.10 The UEQ derived using four components of the object-based PCA method. 63
3.11 The reference layer and the results of the reference layer higher than the mean. (**a**) The reference layer; (**b**) results of the reference layer higher than the mean. 64
3.12 The UEQ derived using the GIS overlay method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean. 65
3.13 The UEQ derived using the pixel-based method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean. 66
3.14 The UEQ derived using the object-based method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean. 66
3.15 The UEQ validation. 67

4.1 The overall workflow for improving the accuracy of UEQ assessment. . . 77
4.2 (**a**) NDVI image derived from the Landsat image (raster data); (**b**) NDVI map after transformation (vector data); (**c**) population layer at the census tract level; (**d**) population layer after transformation to the sub-neighbour level. 78
4.3 (**a**) The LST layer in degrees Celsius before ranking the parameter; (**b**) the ranking LST after the normalization. 79
4.4 The GIS polygons of the parameters. 81

4.5 Weighted distance method. (**a**) k-nearest neighbour; (**b**) Gaussian shape kernel. 84
4.6 The Urban Environmental Quality (UEQ) derived using the GIS overlay method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa. 87
4.7 The UEQ derived using the PCA method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa. 91
4.8 The UEQ derived using the ordinary GWR method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa. 92
4.9 The UEQ derived using the GWR with spatial lag method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa. 93
4.10 The UEQ derived using the GWR with spatial error method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa. 94
4.11 The UEQ results validation. (**a**) The City of Toronto; (**b**) the City of Ottawa. 95
4.12 The reference layer and the higher than the mean of reference layer: (**a**) the reference layer in the City of Toronto; (**b**) the reference layer higher than the mean in the City of Toronto; (**c**) the reference layer in the City of Ottawa; (**d**) the reference layer higher than the mean in the City of Ottawa. 96
4.13 The UEQ derived using the GIS overlay method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa. 97
4.14 The UEQ derived using the PCA method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa. 98
4.15 The UEQ derived using the ordinary GWR method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa. 99

4.16 The UEQ derived using the GWR with spatial lag method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa. 100
4.17 The UEQ derived using the GWR with spatial error method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa. 102

Chapter 1

Introduction

1.1 Research Motivation

The United Nations estimates that the global population will be progressively increased to a double in the coming 40 years, which may cause an adverse impact on the environment and human life. Such impact may instigate increased water demand, overuse of power and anthropogenic noise, etc. One of the key concerns regarding urban planning is to establish certain development goals, such as the Urban Environmental Quality (UEQ).

The terminology "quality of life" has been continuously discussed in the literature so as to lay a foundation to serve the subsequent quantification of UEQ. Szalai (1980) emphasized that quality of life represents the degree of satisfaction with life and the feeling of well-being, which can be measured by the exogenous and endogenous factors. Diener and Suh (1997) concluded the meaning of quality of life should be related to the satisfaction of life. Raphael *et al.* (1996) further echoed and agreed that quality of life is more tend to be the enjoyable degree of a person toward the principal responsibilities of his/her life. However, Van Kamp *et al.* (2003) described the quality of life by the physical and immaterial equipment such as health, education, justice, work, family, etc.

UEQ is the consequence of the combination of environmental parameters including nature, open space, infrastructure, built environment, physical environment amenities and natural resources, where each parameter has its characteristics and partial quality. UEQ is also defined as an indicator to generically describe urban, environmental and socio-economic condition of an urban area. UEQ can be attributed as a multi-layer concept that comprises physical, spatial, economic and social parameters at different scales (Weng and Quattrochi, 2006). Van Kamp *et al.* (2003) addressed that UEQ is an essential part of the quality of life, which has the basic concept such as health, safety and education in addition to the physical and environmental parameters. Weng and Quattrochi (2006) pointed out that UEQ has the capability to influence many governing aspects, including urban planning, infrastructure management, economic influence, policymaking and social studies. Designing a theoretical framework of UEQ linking to the quality of life is an essential step to understand the urban sustainability and human well-being. Such a framework may help to choose the parameters and the integration techniques to evaluate the multidimensional aspects of UEQ (Van Kamp *et al.*, 2003). These integration techniques are able to assess the current and predict/ estimate the future UEQ, which are desired by the municipal and city planners (Leidelmijer *et al.*, 2002). Thus, the assessment of UEQ can be an efficient tool to provide sufficient information about urban conditions, sustainable development and regional planning (Faisal and Shaker, 2017).

UEQ can be modeled using satellite remote sensing techniques through providing continuous Earth observation images and analyzing multi-temporal and multi-resolution data, which are able to give a clear scenario for visualizing and understanding the land cover, Land Surface Temperature (LST), water conditions and vegetation in urban areas (Fung and Siu, 2000, 2001; Nichol and Lee, 2005; Nichol and Wong, 2006; Nichol *et al.*, 2006). However, it is challenging to predict and model the interrelationship and dependence of all the factors. A few preliminary attempts were found using multi-temporal and multi-resolution data to model UEQ (Green, 1957; Bederman and Hartshorn, 1984; Li and Weng, 2007; Nichol and Wong, 2009). As such, UEQ assessment not only provides more detailed information toward urban conditions, it also serves as an efficient tool in sustainable development and urban planning. Subsequently, many representative studies were found in the literature that demonstrated how to use multi-source data to model and assess the UEQ.

1.1.1 UEQ parameters

Background

Previously, modeling the urban and environmental parameters mainly relied on the natural recourse and the effluent disposal areas (Sarmento, 2000; Lo, 1996). Most of the used parameters were mostly coming from physical and chemical parameters. For instance, the Government of Chile in 1978 generated a number of parameters to control water quality for human use as well as the atmospheric emissions that produced by many emissions sources within the country (Sarmento, 2000). Several countries, including United Kingdom, Japan, USA, Portugal and Argentine replicated what has been done by the Chilean Government in 1978. The UK environment department demonstrated the first sustainable development parameters to assess the UEQ (Smeets and Weterings, 1999). Chemical parameters including nitrate, phosphorus, and pesticides as well as other parameters such incident polluted areas, and water distribution and sewage expenses treatment were assigned as essential to assess the UEQ. In 1980, Nakaguchi (1999) investigated several environmental parameters for Japanese cities. Physical and chemical parameters that recommended in the literature were carefully studied to derive adequate parameters in the Japanese cities. These parameters including Concentration of Paniculate Material in Suspension, Biological Oxygen Demand were used to assess the air and water quality respectively. Furthermore, cleanness, noise, the minimum green areas and health resources utilization were also considered to evaluate the UEQ in the Japanese cities.

In Portugal, the agricultural department (AGRO.GES) developed several agricultural parameters to promote the Portugal agriculture regions. The parameters mainly represent structural aspects, conjectural, social and cultural nature to manage the local development dynamics and assess the risks and timing in these regions. In Argentine, the basic infrastructure utilities were assigned as important parameters to measure the UEQ (Sarmento, 2000). The underground supply contamination, public water supply and the final disposal of sewage were utilized for the UEQ assessment. Moreover, socio-economic indicators including income, education, poverty, health and mortality were also used in that study. Other basic utilities including gas, electric energy, drainage and sewerage, public illumination as well as transport and communication were considered to measure the UEQ in Argentine. Noriega and Soria (1999) highlighted several parameters at the beginning of years 1990s for UEQ assessment and sustainable cities. The parameters included air quality, access to green areas, energy consuming, water and generation of solids wastes, accessibility, public participation, the number of people living under the line of poverty and the subjective indicator of welfare.

Choosing appropriate urban and environmental parameters

Before moving forward for choosing the appropriate urban and environmental parameters, we should have a clear definition of the terminologies to better describe the use of the data set. Parameters can be defined as data that is measured or observed for UEQ calculations. For example, population density, crime rate and poverty level can be considered as parameters. Moreover, some other parameters can be also derived from remote sensing, including Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), built-up areas to assess the UEQ. On the other hand, Indicator or some time considered as a value derived from parameters, which provides information about a phenomenon or any environment areas. These indicators can be used to validate and in the assessment of the final UEQ outcome. For example, some of the socio-economic data are considered in this study as UEQ indicators, these include land value, income and education level (Gabrielsen and Bosch, 2003; Indicators for Sustainable Cities, 2015).

Parameters predominantly are designed to provide information about a particular location for a specific purpose to support authorities and decision making. Choosing the appropriate urban and environmental parameters are essential and challenging at the same time. Generally, it begins by reviewing several case studies of various cities that have different size and different environment. From these case studies, the common parameters in several research areas can be assigned. The parameters have to be understandable and easy to implement, and applicable to cities regardless of location and the size of the city (Dekker *et al.*, 2003). Urban planners and policymakers, on the other hand, assigned a sheer number of parameters frameworks, which vary in their approach to measuring UEQ and their selection of parameters (Zavadskas *et al.*, 2007). Most of the parameters frameworks are valid and representative for UEQ. However, some of the parameters frameworks are built for a particular location. For example, the China Urban Sustainability Index was designed to assess cities ranging from 200,000 people to 20 million people in China. On the other hand, the European Green Capital Award indicator frameworks (Berrini and Lorenzo, 2010) was developed to evaluate the current state of the urban and environmental dimension in European cities. Other parameters frameworks named Global City Indicator Program can be used to assess UEQ in all regions. Global City Indicator mainly covers all aspects of urban life, environment and socio-economic, and it does not measure pollution or air quality. The Global City Indicator broaches the economy sector by implementing unemployment rates/ jobs and economic growth, mainly the Annual GDP growth rate. The environment sector also taken into consideration in this parameters frameworks by fulfilling the green spaces, water quality, the volume of solid waste generated and mobility. Moreover, social sector was also considered in this parameters frameworks by including access to local/ neighborhood services within a short distance, crime rates, measures of income distribution, percentage of social/ affordable/ priority housing, percentage of roadways in good condition, percentage of green space, number of schools, percentage of population with access to water or sewage infrastructure, mortality rate and percentage of population with access to health care services (Indicators for Sustainable Cities, 2015).

The subsequent challenge in choosing parameters frameworks is found in investigating the most important parameters towards sustainable development and UEQ. Scholars consensus that the four main parameters including urban, environmental, economic and social are very vital for UEQ research work (Hiremath *et al.*, 2013). In contrast, some researchers have observed that social and economic indicators are not robust enough to represent UEQ (Lynch *et al.*, 2011). Other research workers found that socio-economic indicators including education level and income are required for UEQ assessment (Adelle and Pallemaerts, 2009). People with more education and income are more likely to support high quality environment (Kahn and Matsusaka, 1997; Kahn, 2002). For example, richer urbanites are more likely to support high quality urban areas and purchase good cars that produce less pollution per kilometer (Kahn, 2007). Consequently, several studies showed that income indicator has high relationship up to 0.91 with GDP, car and house ownership in 158 nations in 1996 (Kahn, 2007; Kahn and Matsusaka, 1997; Kahn, 2002). Education provides the tools for people to access and understand information about how environmental hazards affect their wellbeing. As a result, rising educational level can help increase the awareness of individuals for better quality regions (Becker and Mulligan, 1997). Studies also observed a high correlation between the level of education and voting, since people with high education are more likely aware of public/political issues that may influence their environment quality (Kahn, 2002). Thus, socio-economic indicators are essential to assess the UEQ for any urban areas.

There are more issues regarding choosing parameters, which are data standardization and data availability. Since the primary goal of selecting UEQ parameters is to assess the performance of these parameters to better estimate the UEQ. Therefore, the parameters are needed to be standardized and addressed to be on the same scale. In this manner, the selected parameters can be validated and enhanced to serve UEQ precisely (Yigitcanla and Lönnqvist, 2013). Moreover, standardization can also help understanding of the parameters (Pires *et al.*, 2014). Data availability is another significant issue that needs to be considered when parameters are selected for UEQ assessment. Pires *et al.* (2014) highlighted that unavailable data sources could cause a biased or unreliable estimate for UEQ. Researchers consensus that parameters sets need to be locally relevant to the city or municipality (Campbell, 1999; Camagni, 2002). Another scholars emphasized that indicators with extensive political support were more successful than those proposed by academic institutions or non-government agencies (Hiremath *et al.*, 2013). Mega and Pedersen (1998) suggested that indicators should be clear, understandable and obtainable at a reasonable cost-benefit ratio and must be able to reflect every aspect of urban development.

1.2 Research Problem

By reviewing the existing literatures, it was found that there is a lack of quantitative parameters to assess the UEQ. In addition, there is a paucity of research works that discussed the UEQ parameters that can be used to assess the UEQ. The majority of the scholars mainly utilized Principal Component Analysis (PCA), Geographic Information System (GIS) analysis or Multi-Criteria Evaluation (MCE) techniques to integrate UEQ parameters (Nichol and Wong, 2009; Fobil *et al.*, 2011; Lo, 1996; Rinner, 2007), where there exist certain limitations for all these integration techniques. 1) PCA itself produces unweighted components, which may not highlight the importance of the parameters; 2) PCA does not work properly in nonlinear relationships; and finally, 3) the minimum number of components is indeterminable, (Faisal and Shaker, 2017). GIS overlay method does not consider correlation among parameters nor consider data reduction. MCE is a weighting process that allows decision makers to modify attribute values of the parameters. Regarding the result validation, most of the UEQ studies (Fobil *et al.*, 2011; Rinner, 2007; Moore *et al.*, 2006; Lo, 1996; Liang and Weng, 2011) did not perform any field survey or even result validation, except very few attempts found using e-mail questionnaire or field-based questionnaire (Nichol and Wong, 2009; Rahman *et al.*, 2011). Collecting field data is always ideal, but it is also time consuming and budget dependent. Moreover, these methods can be inaccurate to test the outcomes of UEQ if the data samples being collected are not representative, which may lead to biased results.

1.3 Research Objectives

The primary objectives of this Ph.D. research can be summarized as follows:

1. To examine the relationship of multiple UEQ parameters derived from remote sensing, GIS and socio-economic data.
2. To evaluate some of the existing methods (e.g. linear regression, GIS overlay and PCA) for assessing and integrating multiple UEQ parameters.
3. To propose a new method to weight urban and environmental parameters obtained from different data sources.
4. To develop a new method to validate urban and environmental parameters with socio-economic indicators for UEQ assessment.

To broach the first objective, multiple parameters were examined to investigate the relationship among them. The relationship of the environmental parameters including Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), built-up area and LST were assessed with urban parameters including land use, urban density, and public transportation. The environmental and urban parameters was also evaluated with the socio-economic indicators including real Gross Domestic Product (GDP), employment rate and population.

The second objective examines some of the existing methods including linear regression, GIS overlay and PCA to assess the relationships among multiple parameters as well as UEQ integration. The findings can help to look for optimal parameters for UEQ integration and can act as a reference to serve the subsequent comparison of the newly developed UEQ integration method in objective (3).

The third objective is accomplished by proposing a new method to weight urban and environmental parameters. Three Geographically-Weighted Regression (GWR) techniques were investigated to assess the spatial relationship among the parameters. Two major cities in Canada were used as case study to demonstrate the proposed techniques.

Finally, the research explores the accuracy and precision of the final outcome of UEQ. Three socio-economic indicators including family income, the higher level of education and land value were used as a benchmark to validate the final results.

1.4 Dissertation Workflow

Figure 1.1 represents the overall workflow for this dissertation, which can gradually epitomize the steps from data acquisition, data processing, leading to the final outcome of this research work. The primary data of this dissertation mainly obtained from remote sensing, GIS, and census data. The remote sensing and GIS were used to derive the urban and environmental parameters after normalizing the GIS data and atmospheric correction for remote sensing data. The urban and environmental parameters were first investigated and studied to understand its values if they are corresponding to positive or negative values with respect to UEQ. An individual parameter (built-up areas) was assessed using seven major cities in Canada mainly because there is no clear conceptualization in the previous literatures that justify the relationship of built-up regions with respect to UEQ. Then two cities in Ontario, Canada (City of Toronto and City of Ottawa) were tested to derive the UEQ. Since all the parameters came from a different source of data; therefore, all the derived parameters were normalized to be in one scale using Z-score and linear interpolation. Five different integration methods including GIS overlay, PCA, ordinary GWR, GWR with spatial lag and GWR with spatial error model were investigated to derived the final UEQ outcomes. Then three indicators including family income, education level and land values that obtained from census data were used to validate the final results of UEQ.

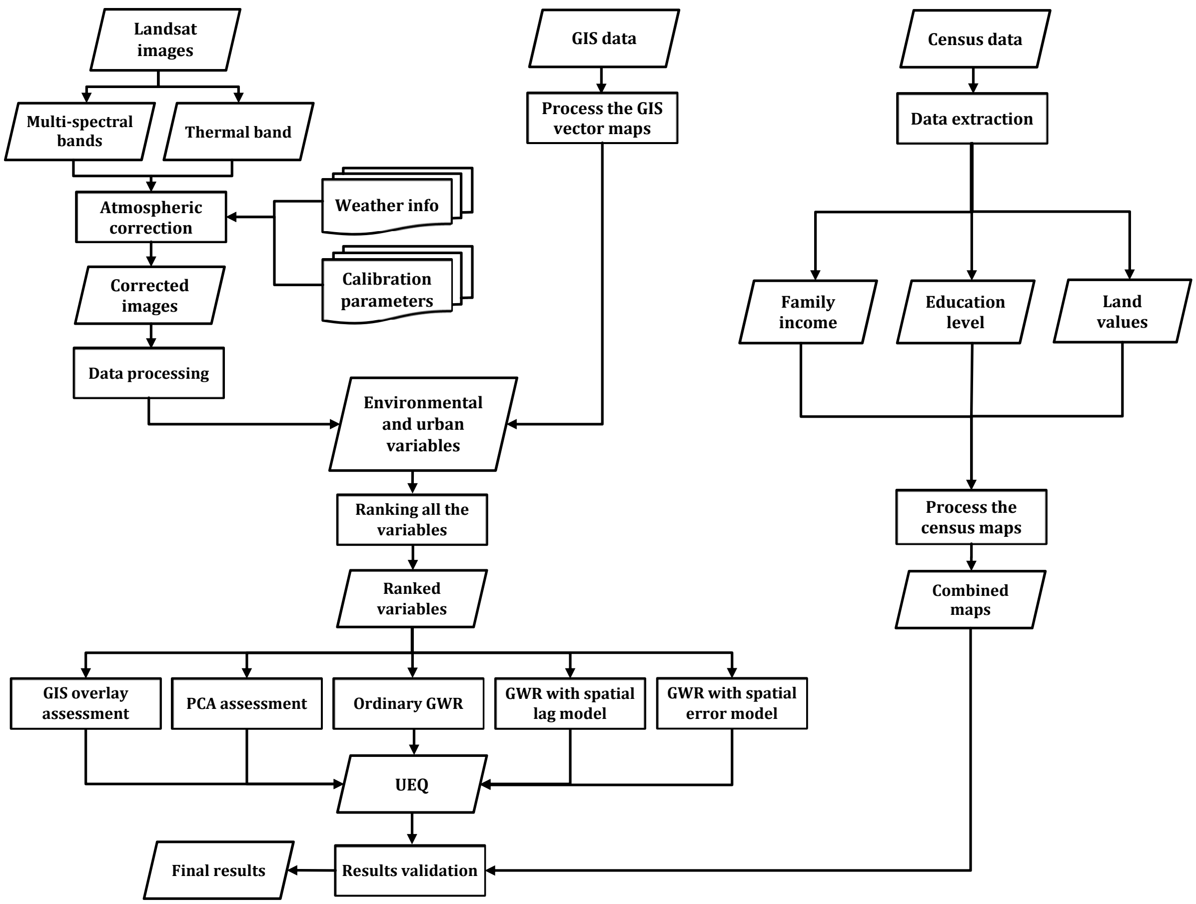


Figure 1.1: The overall dissertation workflow.

1.5 Dissertation Outline

This dissertation follows a manuscript style approach.

Chapter 2 mainly approaches objectives (1) and (2) of the research work. This chapter discusses the relationship between the GDP and built-up area using remote sensing and GIS data. In this chapter, numerous multi-temporal Landsat TM images and land use GIS vector datasets obtained from year 2005 to 2010 during the summer season (June, July and August) for seven major cities in Canada. The socio-economic data, including the real GDP, the total population and the total employment, are obtained from the Metropolitan Housing Outlook during the same period. Both the Normalized Difference Built-up Index (NDBI) and Normalized Difference Vegetation Index (NDVI) were used to determine the built-up areas. This chapter presents the first time use of built-up areas extracted from Remote sensing data in UEQ assessment. Finally, regression analysis was conducted between the real GDP, the total population, and the total employment with respect to the built-up area.This chapter was published as:

Faisal, K., Shaker, A. and Habbani, S. 2016. Modeling the relationship between the Gross Domestic Product (GDP) and built-up area using remote sensing and GIS Data: a case study of seven major cities in Canada. ISPRS International Journal of Geo-Information, 5(3), 23.

In addition, the previous journal paper was broadened from the following conference proceeding:

Faisal, K. and Shaker, A. 2014. The use of remote sensing technique to predict Gross Domestic Product (GDP): an analysis of built-up index and GDP in nine major cities in Canada. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(7), 85.

Chapter 3 contains objectives (1), (2) and (4) of the overall objectives of the research work. This chapter investigates the ability to use remote sensing and GIS techniques to model the UEQ with a case study in the City of Toronto via deriving different environmental, urban parameters and socio-economic indicators. Various remote sensing and GIS data were first explored in order to fully understand the concept of UEQ. The urban, environmental parameters and socio-economic indicators were normalized in this research in order to evaluate the significance of each parameter. GIS overlay and PCA (pixel-based and object-based) were introduced to integrate the urban, environmental parameters and socio-economic indicators with a case study in City of Toronto. Socio-economic indicators, including family income, the degree of education and land value, were used as a reference to validate the outcomes derived from the two integration methods. This chapter was published as:

Faisal K and Shaker A. 2017. An investigation of GIS overlay and PCA techniques for urban environmental quality assessment: a case study in Toronto, Ontario, Canada. Sustainability, 9(3), 380.

In addition, the previous journal paper was broadened from the following conference proceeding:

Faisal K and Shaker A. 2014. Integration of remote sensing, GIS and census data as a tool for urban environmental quality assessment. International Conference on Geospatial Theory, Processing, Modelling and Applications. October 6th to 8th, 2014, Toronto, ON, Canada.

Chapter 4 discusses objectives (3) and (4) of the overall objectives in this research work. This chapter, elucidates the use of the GIS, PCA and Geographically-Weighted Regression (GWR) techniques to integrate various parameters and estimate the UEQ of two major cities in Ontario, Canada. In this chapter, we attempt to fill several gaps in UEQ research by: (1) utilizing a new method to normalize the UEQ parameters; (2) introducing a new approach to weight urban and environmental parameters obtained from diversity data; and (3) proposing a new method to validate urban and environmental parameters with socio-economic indicators for UEQ assessment in two cities in Ontario, Canada. This chapter was published as:

Faisal K and Shaker A. 2017. Improving the accuracy of urban environmental quality assessment using geographically-weighted regression techniques. Sensors, 17(3), 528.

In addition, the previous journal paper was broadened from the following conference proceeding:

Faisal K and Shaker A. 2015. Integration of remote sensing, GIS and census data as a tool for urban environmental quality assessment. IEEE Geoscience and Remote Sensing Society, the International Geoscience and Remote Sensing Symposium 2015 (IGARSS 2015) Milan/ Italy from July 26th to July 31st , 2015.

Chapter 5 epitomized the overall conclusions of the Ph.D. research work along with the limitations and future work. In this chapter, several problems and limitations of the research work are highlighted followed by the overall objectives and the main contribution of this study, which shows how the research responded to the stated objectives. Moreover, this chapter points out the conclusions of the validation and the practical implications of this research.

Chapter 2

Modelling the Relationship between the Gross Domestic Product and Built-Up Area

2.1 Abstract

City/regional authorities are responsible for designing and structuring the urban morphology based on the desired land use activities. One of the key concerns regarding urban planning is to establish certain development goals, such as the real gross domestic product (GDP). In Canada, the gross national income (GNI) mainly relies on the mining and manufacturing industries. In order to estimate the impact of city development, this study aims to utilize remote sensing and Geographic Information System (GIS) techniques to assess the relationship between the built-up area and the reported real GDP of seven major cities in Canada. The objectives of the study are: (1) to investigate the use of regression analysis between the built-up area derived from Landsat images and the industrial area extracted from Geographic Information System (GIS) data; and (2) to study the relationship between the built-up area and the socio-economic data (*i.e.*, real GDP, total population and total employment). The experimental data include 42 multi-temporal Landsat TM images and 42 land use GIS vector datasets obtained from year 2005 to 2010 during the summer season (June, July and August) for seven major cities in Canada. The socio-economic data, including the real GDP, the total population

and the total employment, are obtained from the Metropolitan Housing Outlook during the same period. Both the Normalized Difference Built-up Index (NDBI) and Normalized Difference Vegetation Index (NDVI) were used to determine the built-up areas. Those high built-up values within the industrial areas were acquired for further analysis. Finally, regression analysis was conducted between the real GDP, the total population, and the total employment with respect to the built-up area. Preliminary findings showed a strong linear relationship (R^2= 0.82) between the percentage of built-up area and industrial area within the corresponding city. In addition, a strong linear relationship (R^2= 0.8) was found between the built-up area and socio-economic data. Therefore, the study justifies the use of remote sensing and GIS data to model the socio-economic data (*i.e.*, real GDP, total population and total employment). The research findings can contribute to the federal/municipal authorities and act as a generic indicator for targeting a specific real GDP with respect to industrial areas.

2.2 Introduction

Satellite remote sensors acquire images of the Earth's surface by recording the reflected energy from objects on the ground. Thus, remote sensing data can be used to retrieve semantic information of the Earth surface instead of geometric measurements only. Remote sensing image classification techniques have been used to aid in identifying the land use/land cover areas (Cihlar, 2000; Yan *et al.*, 2015). Land use/land cover features include, without limitation: residential, commercial and industrial, water, vegetation cover and wetland (Selçuk *et al.*, 2003). Applications of remote sensing, particularly in socio-economic studies, aim to map the spatial extent (Huang *et al.*, 2015), urban populations (Sutton *et al.*, 2001), intra-urban population density (Sutton *et al.*, 1997, 2003) and economic activities (Sutton and Costanza, 2002). These data can provide valuable information for the municipal authorities and researchers to aid in urban planning and city management. Real GDP usually serves as an index of the annual production on a country/city's final goods and service. In Canada, the gross national income (GNI) mainly depends on the mining and manufacturing industries and services (Metropolitan Housing Outlook, 2013). Canada is one of the top mining countries, as well as one of the largest producers of minerals and metals. The mining industry contributed 21% of the total value of Canadian goods' exports in 2010, where 28.6% of the industrial sector contributes to the total real GDP in Canada (The Mining Association of Canada, 2011). With respect to the mature development of remote sensing techniques, there emerge several studies using remote sensing data to model the real GDP at a national scale.

Sutton *et al.* (2007) demonstrated a case study to determine the relationships between observed changes in night-time satellite images derived from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) and the changes in total population and real GDP in four nations (India, China, Turkey and the United States). Two approaches were used in their research work to model that relationship. First, the 1992 to 1993 and 2000 DMSP OLS night-time images were used to measure the economic activity within each nation based on the extended light areas. Second, the extended light areas were used to study the relationship with the total population to compute the real GDP, where the state level lights were used to model the relationship with the state level real GDP values based on a linear regression model. However, (Sutton *et al.*, 2007) illustrated that the proposed method is not preferable to measure the real GDP for the developed countries. That is mainly because the night-time image somehow depicts the population density, where the developed countries' GDP may not have a very strong linear relationship with the population density. The results found a strong to moderate positive linear relationship (regression) in this case study (R^2= 0.96 for China, 0.84 for India, 0.95 for Turkey and 0.72 for the United States) and demonstrated a good opportunity to use the remote sensing technique as a tool to map economic activity at national and sub-national levels.

Ma and Xu (2010) conducted a research study in the City of Guangzhou, China. The main goal of the case study was: (1) to detect the urban expansion of the built-up area of the City of Guangzhou in a period of 23 years lasting from 1979 to 2002; (2) to model its urban expansion; (3) to correlate the built-up area of the City of Guangzhou with the real GDP, total population, urban resident income and urban traffic of the city. The supervised image classification technique (maximum likelihood algorithm) was used to classify the image data and extract the built-up areas. Urban expansion was evaluated by analysing the dynamic change of land use. The results showed that the City of Guangzhou extended about 4.5-times from year 1979 to 2002. That gives an indication that the City of Guangzhou expands about 14.2 km^2 on average every year. An optimized trinomial equation was applied to determine the correlation between the built-up areas and socio-economic indicators. The correlation coefficient between the built-up urban area and the total population was found to be about 0.97 within the city. A value of the correlation coefficient of 0.98 was determined between the per capita GDP and the per capita residence area for urban residents in the city (Ma and Xu, 2010).

Ghosh and Elvidge (2010) developed a new tool named the Information and Communication Technology Development Index (IDI) to assess the GDP per capita of different countries of the World. The main goal of the study is to use IDI to measure the development of countries as information societies. The IDI was created by using 11 indicators from information and communication technology (ICT) use, access and skills. The ICT index included three usage indicators (Internet users, fixed broadband and mobile broadband). There are five access indicators, which include in the access index (fixed telephony, mobile telephony, international internet bandwidth, households with computers and households with Internet). The skills index is a very important input for the IDI, because it represents the education within the country. Three indicators (adult literacy, gross secondary and tertiary enrollment) were proposed to represent the skills index in the ICT. Night-time imagery of 2006 was combined with a LandScan population grid of 2006 to measure the human activity within different countries. The LandScan population grid is a method that can be used to determine the total population and the percentage of the total economic activity of the countries. The 2008 World Development Indicators Report provided the per capita GDP record for different countries of the World, where the International Telecommunication Union (ITU) calculated the IDI for the 159 countries. First, the human activity map, which was derived from DMSP-OLS night-time imagery and LandScan, was used to assess the per capita GDP. Second, the per capita GDP map, which was derived from DMSP-OLS night-time imagery and LandScan, were correlated with per capita GDP records obtained from the 2008 World Development Indicators report. The results showed that the relationship reached up to 0.9 R^2 in terms of the regression coefficient. Finally, a second order polynomial regression was used to assess the relationship between the estimated per capita GDP and IDI values, and the results showed about a 0.89 coefficient of regression. The authors demonstrated that remote sensing light images collected at night-time and the LandScan population grid can be used to represent IDI maps and per capita GDP values at finer resolutions (Ghosh and Elvidge, 2010).

Yue *et al.* (2014) proposed another approach in Zhejiang Province located in the southeast China for real GDP estimation. The main objectives of the study is: (1) to propose a low-cost and accurate approach for real GDP estimation by using a diversity source of remote sensing data; and (2) to provide an important database for the government for future developmental strategies. The real GDP was estimated by combining the Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS) night-time imagery, Global MODIS vegetation indices (MODIS EVI), at a resolution of 250 m, and land cover data for 2009. An accurate Human Settlement Index (HSI) was derived by integrating night-time imagery (DMSP/OLS) with the MODIS EVI data in order to estimate the real GDP of secondary and tertiary industries. The land cover data were used to provide the agricultural productivity, such as farming, forestry, stockbreeding and fishery. The land cover data were then used to estimate the real GDP of the primary industries using a threshold mechanism. The brightness values of the night-time imagery (DMSP/OLS) were correlated with the real GDP of secondary and tertiary industries in Zhejiang Province, and the results yielded a correlation coefficient of 0.97. It was found that primary industries, which mainly consist of farming, forestry, stockbreeding and fishery, are hard to detect using DMSP/OLS night-time imagery. That is mainly due to the primary industries only representing 5% of the total real GDP and the coarse resolution of the image.

Despite the above successful attempts, the majority of the studies utilized satellite images collected at night-time that are not always available, since the night-time images are only accessible for National Geophysical Data Centre (NGDC) of National Oceanic and Atmospheric Administration (NOAA) members. The night-time images have low spatial resolution (1 km^2) with respect to the Landsat images, and the night-time imagery (DMSP/OLS) may not be the best option for the estimation of total population and urban areas (Liu *et al.*, 2011). In addition, the above-mentioned studies either focused on the national scale or a specific city. In this research work, the authors aim to utilize remote sensing and GIS techniques to assess the relationship between the built-up area and the reported real GDP in seven major cities in Canada. Instead of using the night-time light images, we employed the multi-temporal Landsat TM satellite images, which are free to the public and have a short revisit time (Yan *et al.*, 2014). As the real GDP of Canada is mainly made up of mining and manufacturing services, thus we place an emphasis on the built-up land with industrial use and compare the remote sensing-derived built-up index with respect to the corresponding real GDP of these seven major cities from 2005 to 2010.

2.3 Datasets and Methods

2.3.1 Datasets

Seven major cities in Canada, *i.e.*, Toronto, Ottawa, Montreal, Québec City, Edmonton, Calgary and Vancouver, were selected in this study due to the data availability and their real GDP differences. The datasets used in this study included three categories of data: (1) Landsat TM satellite images; (2) land use GIS data; and (3) socio-economic data, including real GDP, total population and total employment. Since the land-use GIS data are not available after year 2010, therefore, all of the data were collected from the year 2005 to 2010.

A total of 42 Landsat TM Images were downloaded from the USGS Earth Explorer (United States Geological Survey, 2014). The spatial resolution of the Landsat images is 30 m for the multi-spectral bands and 120 m for the thermal band. All of these images were imported into PCI Geomatics V10.1 (*Geomatica*, version 10.1; PCI Geomatics, Markham, ON, Canada, 2007), an image processing software, clipped and then projected into the UTM coordinate system. Atmospheric correction was conducted with the consideration of the sensor parameters data (sensor type, acquisition date, Sun elevation, Sun zenith and pixel size) and weather conditions (air temperature and visibility). These corrected images were subsequently used to compute the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Building Index (NDBI), as described in Section 2.2. We intentionally ignored those data acquired from November to March due to the appearance of clouds and snow cover, which could affect the experimental results.

A total of 42 land use GIS layer data was acquired from the Scholars GeoPortal (Scholars GeoPortal, 2014) from 2005 to 2010. These land use layers were imported into the ArcGIS environment (*ArcGIS*; Esri; Redlands, CA, USA) for further analysis. Similar to the remote sensing data, all of the data were projected to the corresponding

UTM coordinate system. Socio-economic data are provided by the Metropolitan Housing Outlook (Metropolitan Housing Outlook, 2013) for more than 25 years. The Metropolitan Housing Outlook measures and records the socio-economic indicators, such as the real GDP, total employment and total population, for the major cities in Canada. Table 2.1 summarizes the data sources used in this study.

Table 2.1: The data sources for the seven major cities.

City	**Landsat TM**	**Land Use GIS Data**	**Census Data**
Toronto	Path/Row = 18/30 Date = June to August	Land use vector data were obtained from Scholars GeoPortal in the shapefile format, where the land use categories include: Residential, Commercial, Industrial, Government, Parks, Waterbody and Open Area.	Socio-economic data are provided by the Metropolitan Housing Outlook. Socio-economic data used in this research work include real GDP, total population and total employment.
Ottawa	Path/Row = 16/28 Date = June to August		
Montreal	Path/Row = 15/28 Date = June to August		
Vancouver	Path/Row = 48/26 Date = June to August		
Calgary	Path/Row = 42/24 Date = June to August		
Edmonton	Path/Row = 42/23 Date = June to August		
Québec City	Path/Row = 13/27 Date = June to August		

2.3.2 Methodology

Figure 2.1 shows the overall workflow for this research work, which can be summarized in the following steps. All images were clipped to the study area in each city to speed up the data processing. All image subsets were projected into the UTM coordinate system. Then, atmospheric corrections were carried out on all of the multi-temporal Landsat images. The atmospheric correction model (ATCOR2) developed by (Richter, 1998) was utilized to remove the effects that change the spectral characteristics of the land features (Paolini *et al.*, 2006). To implement the ATCOR2 model, weather information (e.g., air temperature, visibility) were obtained from the Canadian national climate and weather data archive. The calibration parameters for Landsat TM sensor (biases and gains) were also used for atmospheric correction. These calibration parameters are very crucial in the process of atmospheric correction because they provide the bias and gain values in order to convert the image's digital number to radiance and subsequently convert the radiance to top of atmosphere reflectance (Chander *et al.*, 2009). After conducting the atmospheric correction, the bio-physical parameters were derived from the Landsat images. Bands 3, 4 and 5 of the Landsat multi-spectral image were used to determine the bio-physical parameters, NDVI and NDBI, of the study area in order to extract the built-up areas on the images.

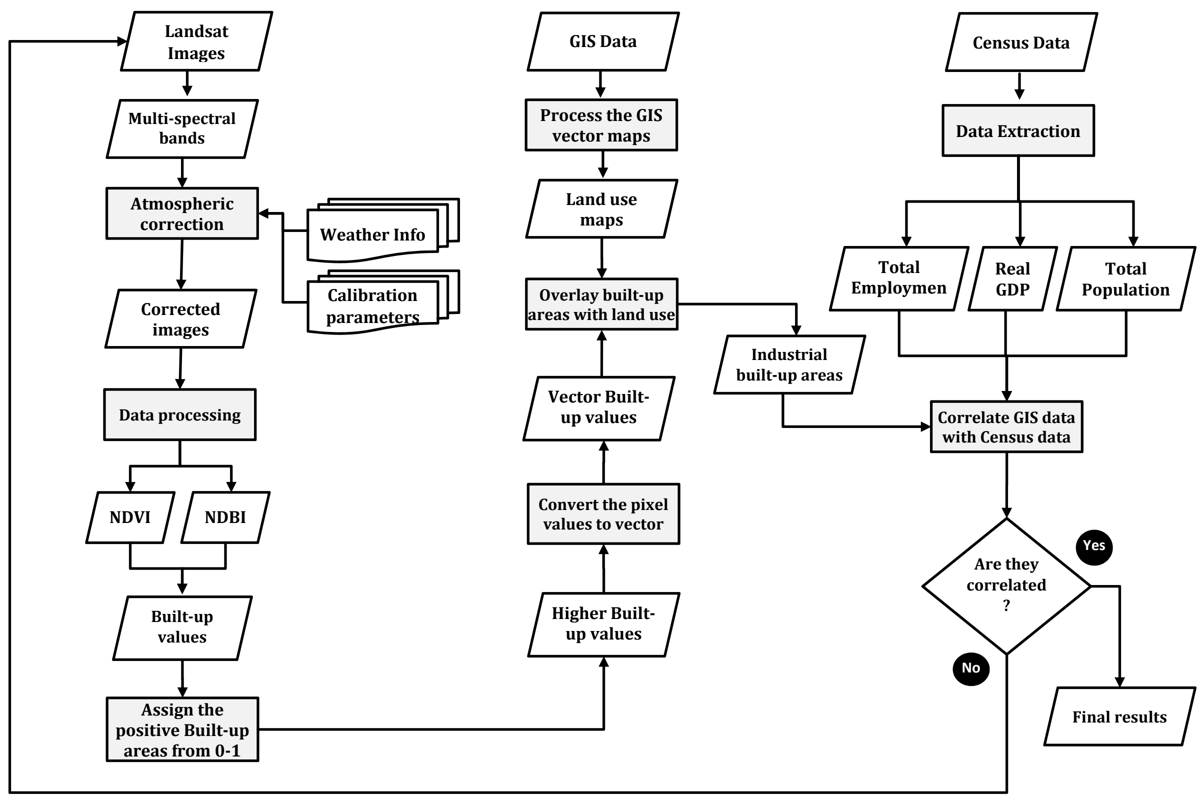


Figure 2.1: The overall workflow for modelling the relationship between the GDP and built-up area.

Regarding how the built-up area being derived, Zha *et al.* (2003) proposed a method to map the urban land (or impervious surface) in the City of Nanjing, China, by using the Landsat TM image due to its high temporal and spectral resolution with respect to other sensors. The cloud-free Landsat TM images were used to derive the NDBI, which represents the built-up regions in the study area. Since some of the vegetation areas were found to be assigned into the built-up category due to the surrounding environments of the vegetation, for that reason, NDVI was calculated to illustrate the vegetation cover in the City of Nanjing. Subsequently, the adjusted built-up areas were derived using arithmetic manipulation between NDBI and NDVI, and only those positive values were classified as built-up areas. Finally, a median filter with a kernel size of 5 pixels by 5 pixels was used to enhance the appearance of the final built-up image. The filtered built-up image was converted to vector data to validate the results using the original colour composite image. It was found that the proposed method yielded an accuracy of 92.6%, which could lead to it being able to be used to map urban areas better than using only NDBI (Zha *et al.*, 2003; Bhatti and Tripathi, 2014; He *et al.*, 2010). Therefore, in this study, we followed the proposed method by Zha *et al.* (2003) to derive the built-up area. First, the NDBI values (ranging from −1 to 1) are calculated using Equation (2.1):

$$\text{NDBI} = \frac{MIR - NIR}{MID + NIR} \tag{2.1}$$

where MIR is the mid-infrared Band 5 of the Landsat TM image and NIR is the near infrared Band 4 of Landsat TM image. The NDVI values (ranging from −1 to 1) refer to an index that is able to monitor the vegetation activity and its annual changes, which can be are calculated using the following equation (Zha *et al.*, 2003):

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red} \tag{2.2}$$

where NIR is the near infrared Band 4 in Landsat image and Red is the red Band 3 in the Landsat image. Finally, the built-up areas are defined by subtracting the NDBI layer from the NDVI layer using the following equation of (Zha *et al.*, 2003):

$$\text{Built-up area} = NDBI - NDVI \tag{2.3}$$

The same concept as in (Zha *et al.*, 2003) is considered, where the positive values obtained from textcolorblackEquation 2.3 represent built-up areas, or otherwise, they refer to non-built-up areas. After that, only those built-up pixels with high positive values were used. Higher positive built-up pixel values were identified based on the histogram of the built-up image. The mean of the histogram of the built-up image was used as a threshold to identify those high values (Faisal and Shaker, 2014). These selected positive built-up pixel values were then converted into polygon shapefile layers for further analysis in ArcGIS (*ArcGIS*; Esri; Redlands, CA, USA). The following Figure 2.2 shows a pictogram to demonstrate how the built-up area is derived.

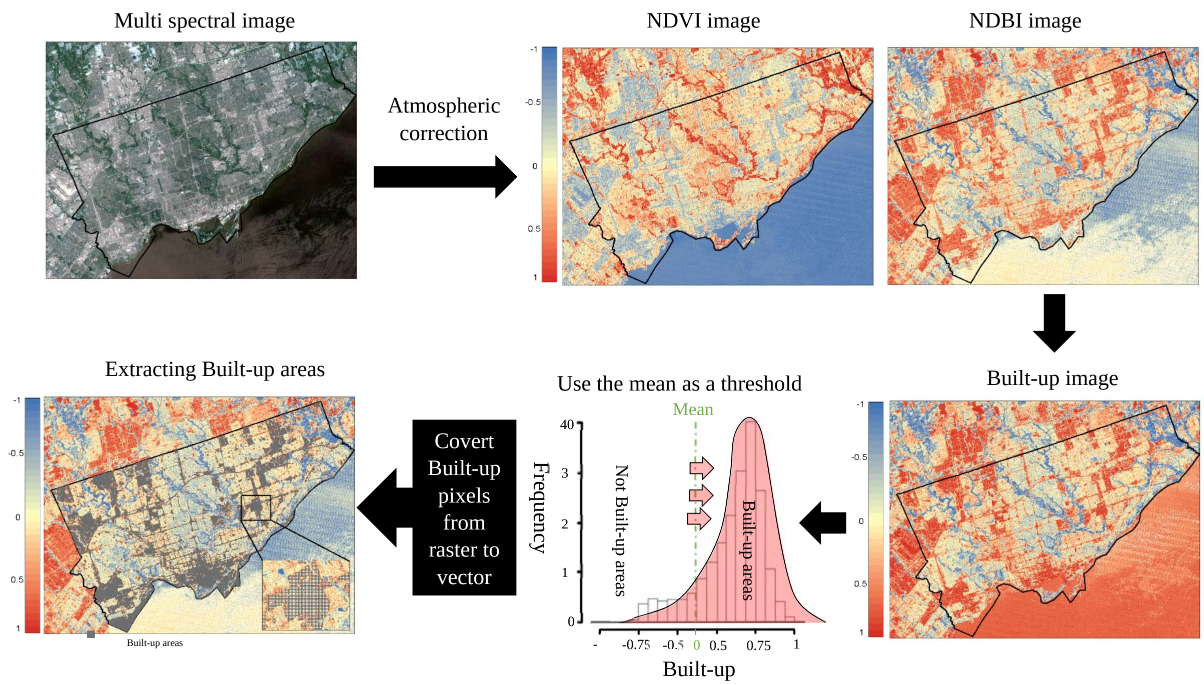


Figure 2.2: A pictogram to demonstrate how the built-up area is derived.

The land use GIS data were used to extract the industrial, commercial and residential areas, and they are used to correlate the Landsat-derived built-up area. All of the GIS data were clipped to the study area in each city to improve the performance of data processing. The built-up pixels were overlapped with the land use maps to calculate the percentage of built-up pixels overlaid on each of the land use areas (industrial, commercial and residential). Figure 2.3 shows an example of the City of Montreal for the built-up image extracted from the Landsat image on the left side of the figure and the land use map extracted from the GIS data on the right side of the figure. The red colour in the built-up image represents the higher built-up values, which reflect buildings or impervious surfaces in the City. The blue colour represents a low built-up value, which also covers some of the vegetation or any green area within the image. Figure 2.4 shows the built-up pixels that were extracted and overlapped with the land use map for the City of Montreal.

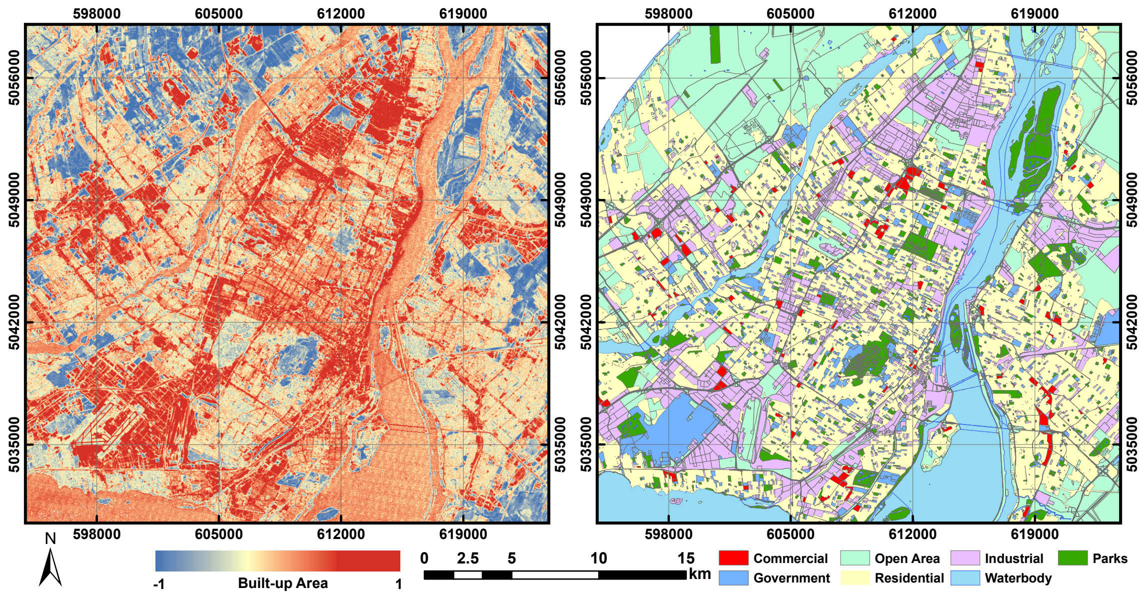


Figure 2.3: Landsat derived built-up image (left) and land use map (right) for the City of Montreal.

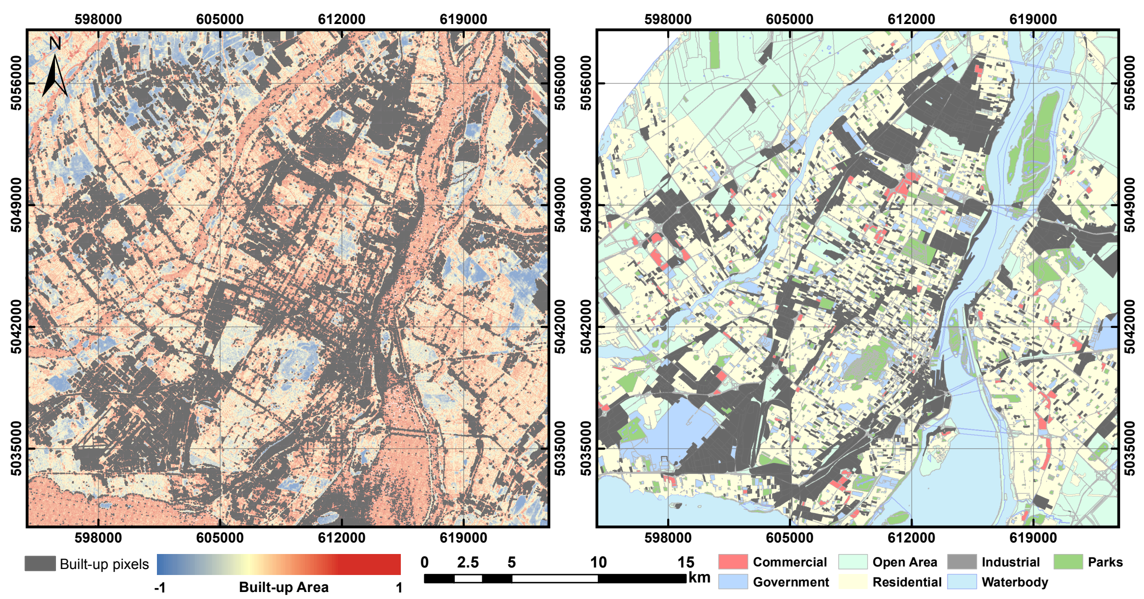


Figure 2.4: An illustration of the built-up pixels overlaid on the land use map.

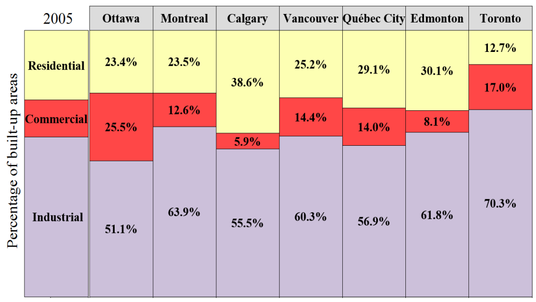
Finally, the social-economic data obtained from the Metropolitan Housing Outlook Metropolitan Housing Outlook (2013), including the real GDP, the total employment and the total population for the seven cities, were plotted and analysed in the GIS platform. Finally, the built-up areas extracted from the Landsat images were correlated with the three socio-economic indicators in order to reveal their relationships. The linear regression analysis was used to depict the relationship between any two parameters, and in this study, we used the R^2, *i.e.*, the coefficient of determination, as an indicator to reveal the relationship between any of these two parameters. A confidence interval of 95% was used throughout the linear regression analysis.

2.4 Results and Discussion

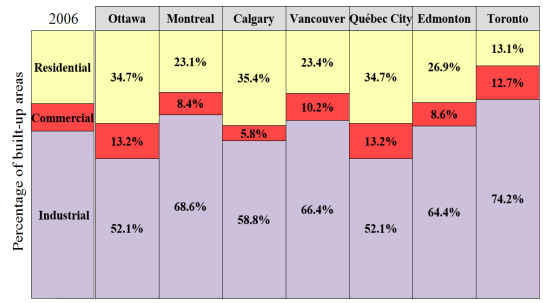
2.4.1 Built-Up Areas

In this section, an analysis was first conducted to reveal the relationship between the Landsat-derived built-up areas and the underneath land cover zones. Figure 2.5 shows the percentage of the built-up areas derived from the Landsat images correlated with the land use maps from the GIS data. The total number of pixels in the built-up area was calculated from each land use (industrial, residential and commercial), where the percentage of the built-up areas was computed for each land use in the map. Most of the built-up areas derived from the Landsat images are located in the industrial and residential zones. However, the built-up areas are mainly occupied in the industrial zones by 51%–70% in 2005, 52%–70% in 2006, 50%–67% in 2007, 45%–73% in 2008, 51%–75% in 2009 and 51%–80% in 2010. That is mainly because of the higher reflectance of the industrial areas that are usually paved with homogeneous concrete and asphalt structures compared to the residential and commercial areas. The industrial areas are mainly covered by a large extent of concrete structure without any distinguished vegetation on site, where the residential areas contain residential buildings and houses. Many of these residential buildings and houses have vegetation surrounding them, which may influence their corresponding spectral reflectance values found within the Landsat images.

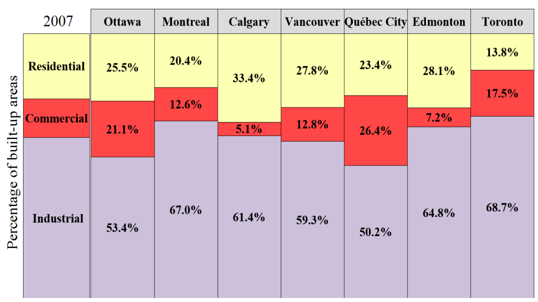
In 2005, the built-up areas obtained from the Landsat images are consistently located within the industrial zones in the seven cities. The percentage of the built-up areas found within big cities, such as the City of Toronto, Montreal and Vancouver, has a higher



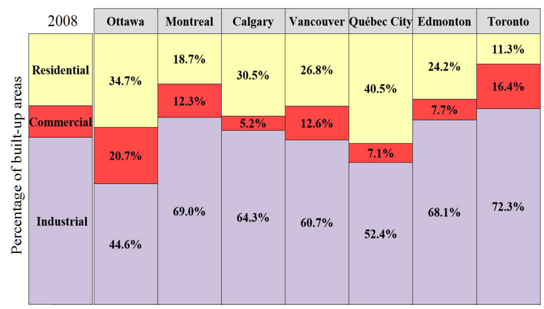
(a)



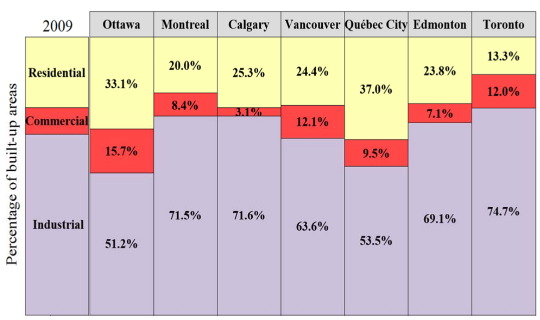
(b)



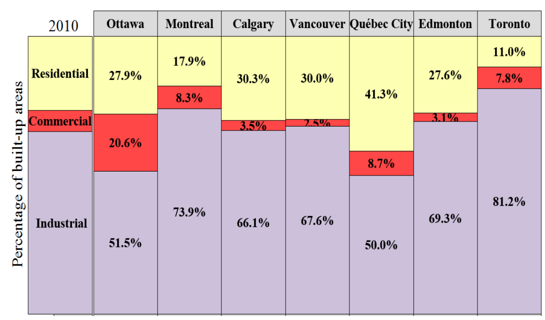
(c)



(d)



(e)



(f)

Figure 2.5: Percentage of built-up areas within the land use. (**a**) 2005; (**b**) 2006; (**c**) 2007; (**d**) 2008; (**e**) 2009; (**f**) 2010.

percentage of the built-up areas compared to small cities, such as Québec City, by 10% to 20%. However, the percentage of the built-up areas within big cities, such as Toronto, Montreal and Vancouver, is significantly higher than the percentage of the built-up areas in small cities, such as Québec City, by 30% to 36% for the year 2010. Such findings can be explained due to the fact that the industrial areas in those big cities occupy more land than those industrial areas in the small cities by 70 to 100 km^2. The highest percentage of the built-up areas within the industrial zones in 2010 is found in the City of Toronto (81%). The lowest percentage of the built-up areas within the industrial zones is located in Québec City (50%). This could be due to the variation of the manufacturing and services industries in Toronto, which are found to be more significant than that in Québec City.

Figure 2.5 shows dramatic changes in the percentage of the built-up areas that are located within the residential and commercial areas in the cities. The percentage of the built-up areas that are located within the residential and commercial areas in Québec City and Ottawa changed from 4.5% to 12% throughout 2005 to 2010, which could be explained as due to the urban sprawl in those two cities. However, the industrial areas in the other cities, such as Toronto, Montreal, Vancouver and the City of Calgary, expanded by 7% to 10% from the year 2005 to 2010. This can be explained by the urban expansion in the industrial sector that is more than the residential and commercial sector in these four cities.

The highest percentage of the built-up areas within the industrial zones in 2005 is located in the City of Toronto (70%). The lowest percentage of the built-up areas within the industrial zones is located in the City of Ottawa (51%). That is mainly because of the urban sprawl, where the total population in the City of Toronto is about five million people. However, the combined total population of Québec City and the City of Ottawa is about two million people in 2005 (Metropolitan Housing Outlook, 2013). From the year 2005 to 2010, it was observed that the built-up areas, which are located within the industrial zones, are significantly higher than the built-up areas, which are located within the residential and commercial areas zones. For that reason, further analysis was conducted to determine the linear regression between the percentage of the built-up areas and the industrial areas from year 2005 to 2010. A strong positive linear relationship was observed for all of the built-up areas and industrial areas, where R^2= 0.82 is detected for the percentage of the built-up areas found within the industrial zones from 2005 to 2010, as shown in Figure 2.6. With these findings, one can conclude that the Landsat-derived built-up areas mainly represent the industrial zones regardless of the cities being analysed. This thus paves the way for the subsequent analysis, in the following Section 2.4.2 which aims to model the relationship between the Landsat-derived built-up areas with respect to the real GDP, the total population and the total employment.

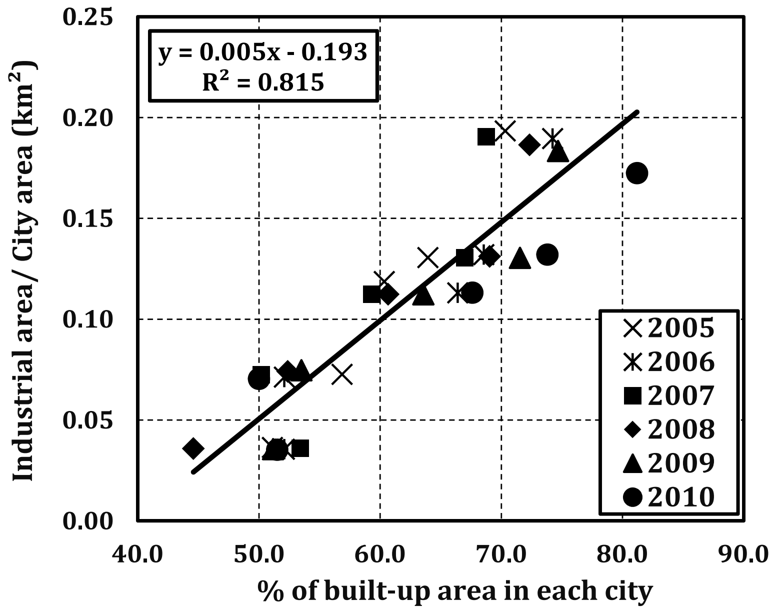
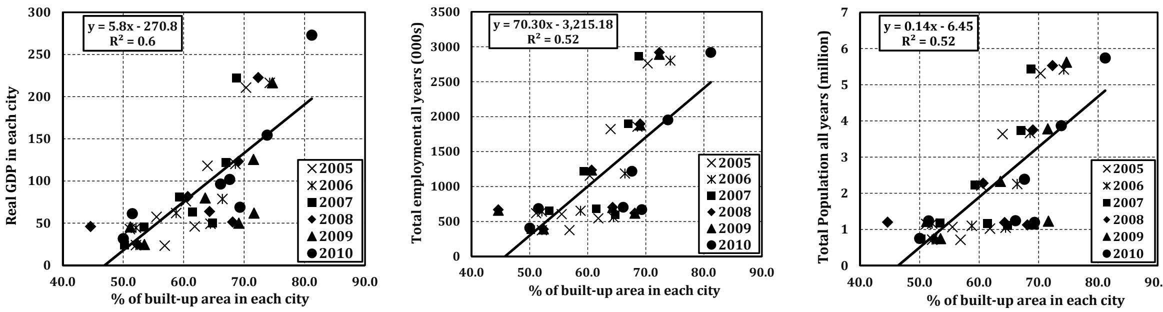


Figure 2.6: Relationship between the percentage of built-up areas and the industrial areas from 2005 to 2010.

2.4.2 Regression Analysis between the Socio-Economic indicators and Built-Up Areas

A preliminary analysis was conducted to determine the linear regression between the real GDP, total employment and total population from socio-economic indicators with respect to the percentage of the built-up areas derived from the remote sensing images within the industrial zones from the year 2005 to 2010. Such analyses are found missing in the existing literature, which adopted Landsat images for industrial land use and socio-economic indicators.

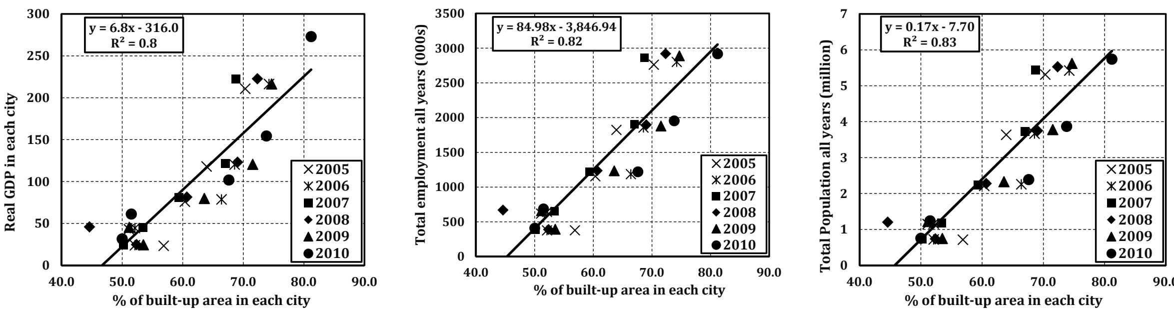
Figure 2.7 shows the relationship between the percentage of the built-up areas within the industrial zones and the real GDP, total employment and total population from 2005 to 2010, respectively. Preliminary analysis revealed that a moderate positive linear relationship exists for both of the socio-economic indicators and the percentage of the built-up areas within the industrial zones from 2005 to 2010. An R^2 of 0.6 was detected for the percentage of the built-up areas within the industrial zones and real GDP. On the other hand, an R^2 of 0.5 was observed for the percentage of the built-up areas within the industrial zones and total population. Furthermore, the linear regression between the percentage of the built-up areas found within the industrial zones and R^2 of total employment was 0.5.



(**a**) (**b**) (**c**)

Figure 2.7: Relationship between % of built-up areas and the socio-economic indicators from 2005 to 2010. (**a**) % of built-up areas *vs.* GDP; (**b**) % of built-up areas *vs.* total employment; (**c**) % of built-up areas *vs.* population.

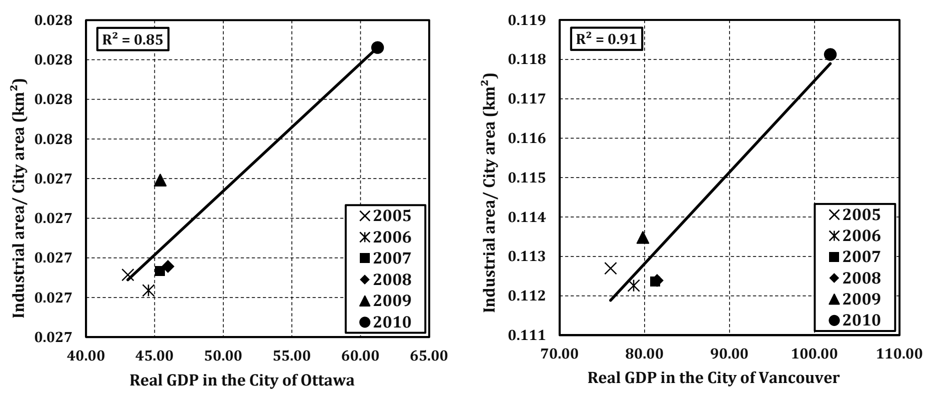
Since a few outliers were observed in Figure 2.7, which are mainly contributed by the City of Calgary and Edmonton, if the data of these two cities were eliminated from the analysis, the linear regression between the socio-economic indicators and the percentage of the built-up areas within each city were significantly improved. The R^2 between the percentage of the built-up areas within the industrial zones and the real GDP jumped from 0.6 to 0.8. The linear regression between the percentage of the built-up areas within the industrial zones and total population increased from 0.5 to 0.83. The regression between the percentage of the built-up areas within the industrial zones and total employment improved from 0.5 to 0.82, as shown in Figure 2.8. The reason for deducting these two cities for analysis is mainly because the City of Edmonton and the City of Calgary are located in the province of Alberta, which is mainly dependent on the oil and gas industries (Canadian Centre for Energy Information, 2012), where most of the oil and gas manufacturers are located outside of the cities. For that reason, the industrial areas within the cities may not accurately represent the real GDP of the cities. As a result, the cities of Calgary and Edmonton were eliminated from the data due to the low regression between industrial areas and socio-economic indicators.



(**a**) (**b**) (**c**)

Figure 2.8: Relationship between % of built-up areas and the socio-economic indicators without Edmonton and Calgary from 2005 to 2010. (**a**) % of built-up areas *vs.* GDP; (**b**) % of built-up areas *vs.* total employment; (**c**) % of built-up areas *vs.* population.

In spite of these results, all of the fitted regression lines show that the percentage of the built-up areas within the industrial zones has a direct proportional relationship to all of the socio-economic indicators that were used in this study. Further analysis was conducted to determine the linear regression between the industrial area from GIS data and the real GDP from 2005 to 2010. That is mainly to investigate which city inflates the overall regression. As noted in the below Figures 2.9 and 2.10, the results vary in each of the cities.



(**a**) (**b**)

Figure 2.9: Relationship between the real GDP and industrial areas from 2005 to 2010. (**a**) City of Ottawa; (**b**) City of Vancouver.

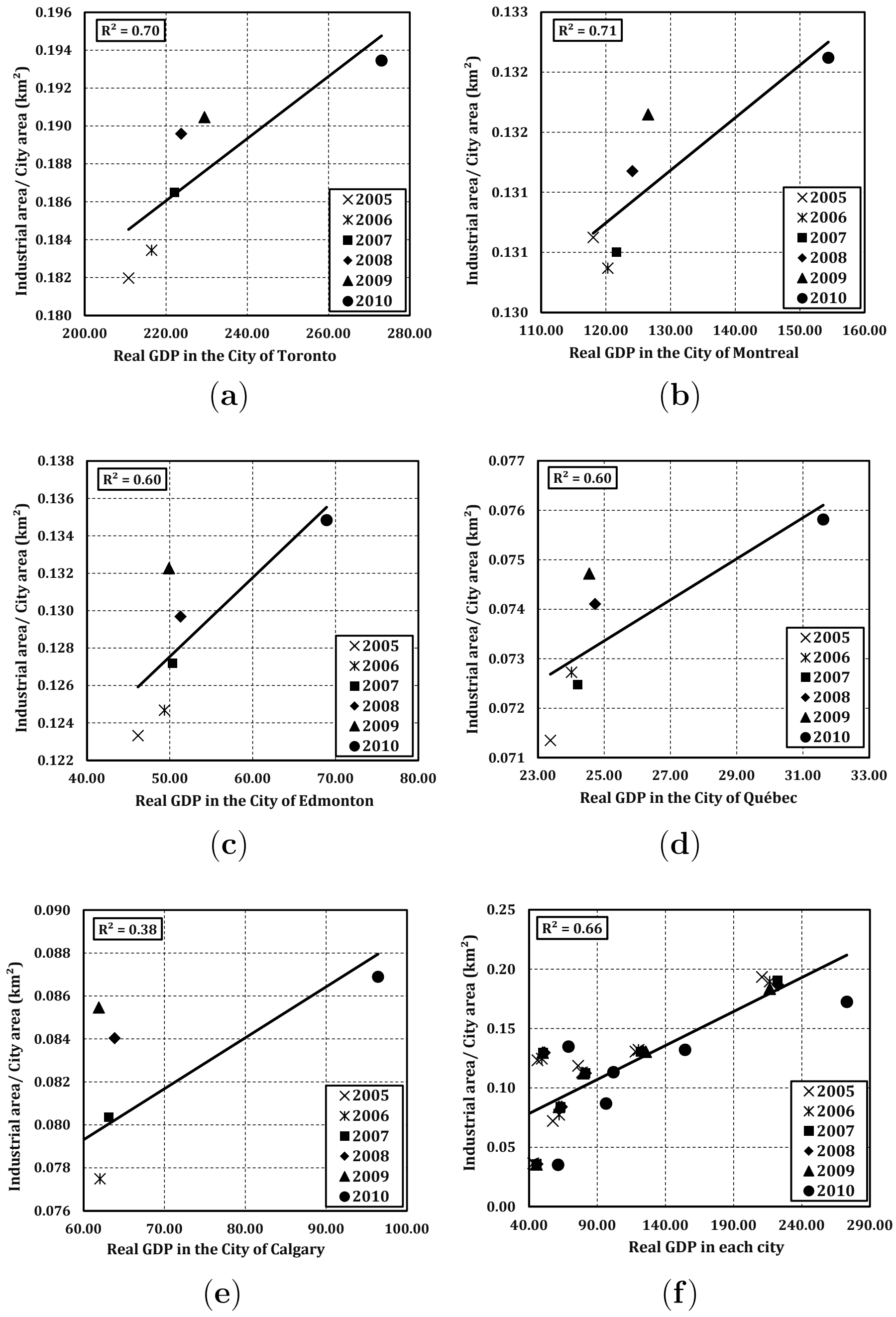


Figure 2.10: Relationship between the real GDP and industrial areas from 2005 to 2010. (**a**) City of Toronto; (**b**) City of Montreal; (**c**) City of Edmonton; (**d**) Québec City; (**e**) City of Calgary; (**f**) all cities.

A strong positive linear relationship was observed for both of the industrial areas and real GDP in the City of Ottawa and the City of Vancouver, resulting in a R^2 of 0.9 and 0.8 from 2005 to 2010, as shown in Figure 2.9 a,b. A moderate positive linear relationship (R^2= 0.7 and 0.6) was found for the cities of Toronto, Montreal, Edmonton and Québec City, as shown in Figure 2.10 a to d. The City of Calgary has a weak linear relationship ($R^2 = 0.4$) between the industrial areas within the cities and the corresponding GDP from 2005 to 2010, as shown in Figure 2.10 e.

Based on the previous observation, the City of Edmonton and Calgary have negative impact on the overall regression, because these two cities are mainly dependent on the oil and gas industries (Canadian Centre for Energy Information, 2012), where most of the oil and gas manufacturers are located outside the cities, as mentioned previously. Therefore, a moderate positive linear relationship (R^2= 0.66) was determined, when Edmonton and Calgary were involved in the dataset, as shown in Figure 2.10 f. However, a strong positive linear relationship was observed for both of the industrial areas and real GDP if the aforementioned two cities were eliminated, resulting in a R^2 of 0.81 for the real GDP from 2005 to 2010, as shown in Figure 2.11.

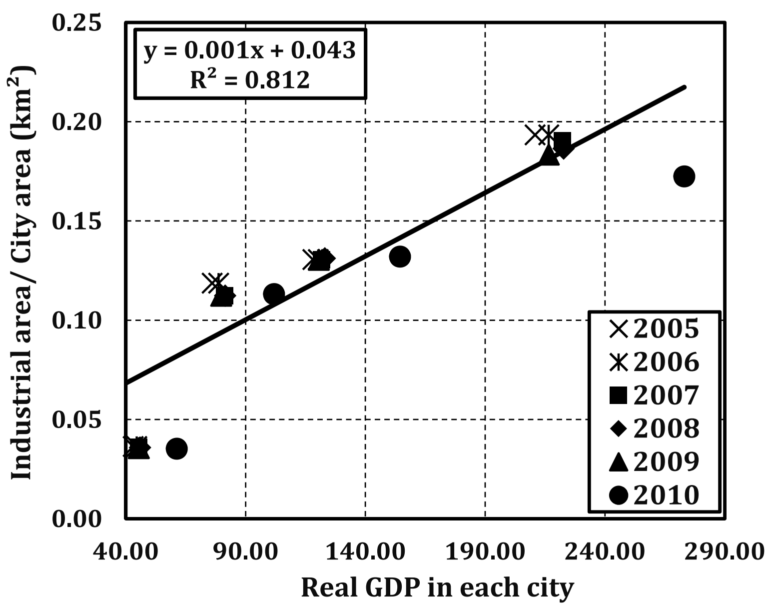


Figure 2.11: Real GDP *vs.* industrial areas without Edmonton and Calgary.

Despite these moderate/strong regressions being revealed, such a method may not be replicated at the individual city level, because remote sensing-derived parameters are unable to explain the large amounts of variance in GDP. Current economic development studies have already pointed out certain factors influencing the GDP, including energy consumption, foreign direct investment and CO_2 emissions (Pao and Tsai, 2011). In addition, some parameters that contribute to the GDP, such as retail sales, service sector and manufacturers' shipments, are hard to measure (Landefeld *et al.*, 2008). Therefore, the use of the remote sensing technique to model the GDP only contributes to a certain degree (in a particular spatial dimension), while other socio-economic and environmental factors should be considered in order to derive a more universal indicator to predict the economic development at the country-wide level. Therefore, all of these hidden factors may affect the regression coefficient (R^2) in each city.

2.4.3 Discussion

In summary, this study aims to investigate the ability of using remote sensing technique to model and predict the real GDP for those cities that are mainly dependent on industrial and manufacturing incomes. To achieve this, we have to first prove there exists a relationship between the remote sensing-derived indices (*i.e.*, the built-up areas) with respect to the industrial zones, which has been reported in Section 2.4.1 With such a high linear regression (R^2= 0.82) between the remote sensing-derived built-up areas, as well as the industrial areas, one can assume that the built-up areas have a high component of industrial and manufacturing activities. Therefore, in Section 2.4.2, we investigate the relationship between the Landsat-derived built-up areas with respect to the real GDP, the total population and the total employment. A high linear regression was observed ($R^2 = 0.8$) for the three socio-economic indicators. Thus, the presented approach can be replicated by any federal authorities in developed countries, where their major incomes are dependent on the industrial and manufacturing activities. However, there are a few varieties of limitations in regards to the research study. (1) More up-to-date GIS data are required to consolidate the findings for different cities and countries. (2) Water bodies and bare soil all have high built-up index values that may cause confusion with the impervious surfaces. If this method were being applied elsewhere and no GIS data exists, it would likely cause problems, as water bodies or bare soil could be classified as built-up areas, and the relationship with GDP would be affected. (3) Other regression analyses (such as nonlinear regression) can be explored depending on the nature of study area and the socio-economic indicators being studied. The authors have investigated the use of nonlinear regression to run the relationship between the GIS and remote sensing data, with respect to the socio-economic data, including real GDP, total population and total employment. However, there is no consistent trend being found in all of these cities regardless of the improvement on R^2 that thus reveals the inappropriate use of a non-linear model in this specific case study.

Such an argument is somehow supported in the existing literature (Sutton *et al.*, 2007) and (Yue *et al.*, 2014), where all of these studies examined the use of the linear regression approach to analyse the data, which either focused on the national scale or on a specific city, and the majority of the results represent moderate to strong linear regression between the remote sensing-derived information with respect to the socio-economic data. Although the indicators being analysed may not be identical, the use of linear regression somehow has its grounds in accordance with these existing literatures. In short, the remote sensing technique can provide fruitful information to model some of the socio-economic indicators. However, other socio-economic indicators and empirical models should be considered in order to develop a more universal indicator to predict the GDP. As a result, further research should be carried out to reveal the relationship with respect to other parameters, such as energy consumption, foreign direct investment and CO_2 emissions (Pao and Tsai, 2011), as well as developing a new technique to retrieve the built-up area for those regions located in an arid environment and cold region or specifically designed city, like a green city.

2.5 Conclusions

This study aims to investigate the relationship between the built-up area, as well as three socio-economic indicators (the real GDP, total population and total employment) in order to facilitate any new city development and regional planning, with a case study of seven major cities in Canada. Since not all of the cities have a comprehensive GIS land use dataset to find out the built-up areas, thus we proposed to utilize remote sensing data to estimate the built-up areas to achieve such a goal. In this study, we analysed 42 Landsat images and 42 land use maps in order to study the regression between the percentage of built-up areas extracted from the satellite image and the reported real GDP in seven major cities in Canada. The Landsat TM images were first atmospherically corrected, and the built-up values were computed using the NDBI and NDVI. Those high built-up values within the industrial areas were derived from the Landsat images for subsequent analysis. Built-up values within the industrial areas were correlated with industrial zones within the seven cities with a strong positive linear relationship ($R^2 = 0.82$) found from the year 2005 to 2010.

A further analysis was conducted to investigate the regression between the real GDP, population and total employment with respect to the built-up areas. It was found that the percentage of built-up areas, which are located in the industrial zones, has a moderate positive linear relationship ($R^2 = 0.6$, and 0.5) with the socio-economic indicators if all cites were considered in the datasets. However, an improvement of the coefficient regression ($R^2 = 0.8$, 0.82 and 0.83) was observed when the City of Edmonton and Calgary were eliminated from the analysis, since these cities have a relative high gross income from the oil mining industry that does not require a large piece of land for manufacturing. With the regression found, the results can be used as a generic indication for the federal/municipal authorities, which are aiming at or targeting a specific real GDP with respect to the planned industrial areas for city management. Future work can be focused on developing a new method to accurately extract the built-up area for arid or cold region environment/country, since the NDBI-NDVI approach may not be applicable in those area to extract the built-up zone. In addition, the relationship between the real GDP, as well as other remote sensing-derived indices (such as area of desert or soil) should be investigated for those regions.

Chapter 3

An Investigation of GIS Overlay and PCA Techniques for UEQ Assessment: A Case Study in Toronto, Ontario, Canada

3.1 Abstract

The United Nations estimates that the global population is going to be double in the coming 40 years, which may cause a negative impact on the environment and human life. Such an impact may instigate increased water demand, overuse of power, anthropogenic noise, etc. Thus, modelling the Urban Environmental Quality (UEQ) becomes indispensable for a better city planning and an efficient urban sprawl control. This study aims to investigate the ability of using remote sensing and Geographic Information System (GIS) techniques to model the UEQ with a case study in the city of Toronto via deriving different environmental, urban parameters and socio-economic indicators. Remote sensing, GIS and census data were first obtained to derive environmental, urban parameters and socio-economic indicators. Two techniques, GIS overlay and Principal Component Analysis (PCA), were used to integrate all of these environmental, urban parameters and socio-economic indicators. Socio-economic indicators including family income, higher education and land value were used as a reference to assess the outcomes derived from the

two integration methods. The outcomes were assessed through evaluating the relationship between the extracted UEQ results and the reference layers. Preliminary findings showed that the GIS overlay represents a better precision and accuracy (71% and 65%), respectively, comparing to the PCA technique. The outcomes of the research can serve as a generic indicator to help the authority for better city planning with consideration of all possible social, environmental and urban requirements or constraints.

3.2 Introduction

Urban Environmental Quality (UEQ) is defined as an indicator to generically describe the urban, environmental and socio-economic condition of an urban area. UEQ can be regarded as a multi-layer concept that comprises physical, spatial, economic and social parameters at different scales (Weng and Quattrochi, 2006). Weng and Quattrochi (2006) addressed that UEQ has the capability to influence many governing aspects, including urban planning, infrastructure management, economic influence, policy-making and social studies. However, it is challenging to predict and model the inter-relationship and dependence of all of the factors. Recently, satellite remote sensing techniques can help in modelling UEQ through providing continuous Earth observation images of the urban environment at different spatial, spectral and temporal resolutions (Nichol and Lee, 2005; Nichol and Wong, 2006; Nichol *et al.*, 2006). A few preliminary attempts were found using multi-temporal and multi-resolution data to model UEQ (Green, 1957; Bederman and Hartshorn, 1984; Li and Weng, 2007; Nichol and Wong, 2009), since these data can provide a clear vision for visualizing and understanding the land cover, water conditions and vegetation in urban areas (Fung and Siu, 2000, 2001). As such, UEQ assessment not only provides more detailed information toward urban conditions, it also serves as an efficient tool in sustainable development and urban planning. Subsequently, a number of representative studies were found in the literature that demonstrated how to use multi-source data to model and assess the UEQ.

Nichol and Wong (2009) conducted a research study in the Kowloon Peninsula, Hong Kong. The main goal of the study was to investigate different ways of combining six parameters (vegetation density, heat island intensity, aerosol optical depth, building density, building height and noise) in different units into a single integrated UEQ index and to establish a suitable mapping scale at which these parameters operated and interacted. The study was conducted at two scale levels through using a high resolution IKONOS satellite image and a fine resolution Landsat satellite image. Two approaches, including Geographic Information System (GIS) overlay analysis and Principal Component Analysis (PCA), were used to integrate these six parameters. To act as a reference of UEQ assessment, an email questionnaire survey was conducted with 200 Kowloon Peninsula residents to weight the parameters. The field-based questionnaire survey was administered at 70 locations during the summer season to validate the results. The results showed that the combined parameters, including vegetation density, building density and building height, are more representative to model the UEQ in Hong Kong. Moreover, the overall result showed that the UEQ result derived by GIS overlay analysis is deemed to be close to the residents' opinion obtained from the questionnaire surveys.

Liang and Weng (2011) introduced various environmental parameters and socio-economic parameters to assess UEQ changes in Indianapolis, USA, in the past 10 years. A total of 18 environmental parameters, including cropland and pasture, water, forest, grass, barren lands, commercial and industrial areas, high density residential areas, medium density residential areas, low density residential areas, Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Normalized Difference Built-up Index (NDBI) from transformed bands, were extracted from two Landsat Thematic Mapper (TM) images taken in 1991 and 2000; among which, 13 socio-economic indicators, including population density, median age population, household, house unit, owner-occupied house unit, vacant house unit, median house income, median family income, per capita income, house value, percentage of college graduate, percentage of family under poverty line and unemployment rate, were derived from US census of 1990 and 2000. The results demonstrated that four principal components being extracted can sufficiently represent the 28 parameters. The variance of each component was used as a weight to compute the UEQ for each year. The derived UEQ map showed that medium UEQ zone in 1990 was recognized as poor zones in the 2000. High UEQ areas in 1990 were transformed to medium UEQ areas in 2000. The city centre in 1990 was identified as a mixture of high, medium and poor UEQ. However, in 2000, the city centre became a medium level UEQ zone. The UEQ was significantly improved in the south and the southeast zones over the past ten years in Indianapolis, USA.

Another representative study covered the city of Delhi, India, conducted by (Rahman *et al.*, 2011). The city has been suffering from a dramatic increase of population annually, which has led to environmental and public services degradation. The east district of Delhi had the largest population with 98.75% of the urban population in 2001. The main goal of this case study was to investigate the UEQ in the east district of Delhi. The Advanced Space-borne Thermal Emission Reflection Radiometer (ASTER) image of the year 2003 was obtained to generate the land use/land cover map. The guide map was used to generate the land use/land cover map of year 1982. Supervised classification was conducted based on the maximum likelihood method. Five land classes (named high density residential, medium density residential, low density residential, roads and open green spaces) were extracted from the ASTER image. Socio-economic indicators and environmental parameters, such as built-up area, open spaces, household density, occupancy ratio, population density, accessibility to roads, noise and smell affected area, were used to assess the UEQ of Delhi. GIS overlay was conducted to integrate the urban environmental parameters for the years 1982 and 2003. The result showed that in the year 1982, 89% of the east district of Delhi was in good environmental conditions, while the remaining areas were in fair conditions or bad alarming conditions. However, in the year 2003, 75% of the east district of Delhi was in good environmental condition, while the areas with poor and bad alarming conditions had increased to 22% and 3.5%, respectively. The reason is mainly due to the unplanned urban extension found within the east district of Delhi. The study demonstrated that remote sensing and GIS data are viable techniques for urban environmental management and decision making.

Rinner (2007) investigated the combination of Geographic Visualization (GeoVis) and Multi-Criteria Evaluation (MCE) methods to assess the UEQ within Toronto. MCE is a weighting method that allows decision makers to modify attribute values of the parameters. Numerous socio-economic and demographic indicators, including population change, population density, ownership of dwellings, family size, average household income, expenditure on housing, employment status, immigration status and degree of education were used to assess the UEQ. An analytic method, named the Analytic Hierarchy Process (AHP), was investigated to estimate the composite measures of quality of life. The AHP method can be used to visualize the spatial patterns and combine different models for

UEQ. Interviews with three senior geography students were conducted to validate the results. The result of this case study is more toward supporting analysts to review their final decision-making strategies.

Despite the above successful attempts, the majority of the UEQ studies utilized PCA or GIS analysis techniques to integrate various parameters (Nichol and Wong, 2009; Li and Weng, 2007; Liang and Weng, 2011; Rahman *et al.*, 2011). Although PCA is an analytical technique that compresses the main data into lower dimensions that retain most of the data variance (Jensen, 2005), the method still has several potential drawbacks: (1) it produces unweighted components, which may not represent meaningful parameters; (2) PCA does not work properly in nonlinear relationships; and finally, (3) the minimum number of components is indeterminable. Although some researchers used the GIS overlay method to integrate different parameters (Nichol and Wong, 2009; Rahman *et al.*, 2011), the GIS overlay method does not consider the correlation among the parameters. Each parameter may rank from a certain range, say 1–10, where 10 represents the best condition and 1 represents the worst condition. The sum of the derived parameters corresponds to the UEQ ranking. The GIS overlay method can be used effectively to store, analyse and represent layers from different types of map features (Nichol and Wong, 2009). Regarding the result validation, most of the UEQ studies (Fobil *et al.*, 2011; Rinner, 2007; Moore *et al.*, 2006; Lo, 1996; Liang and Weng, 2011) did not perform any field survey or even result validation, except very few attempts found using e-mail questionnaire or field-based questionnaires (Nichol and Wong, 2009; Rahman *et al.*, 2011). Collecting field data is always ideal, but it is also time consuming and budget dependent. Moreover, these methods can be inaccurate to test the outcomes of UEQ if the data samples being collected are not representative, which may lead to bias results.

In this research, the main objectives are: (1) to investigate GIS overlay and PCA techniques to assess UEQ with a case study in the city of Toronto, Ontario, Canada; (2) to test a new approach to normalize the data derived from remote sensing and GIS data; and (3) to assess a new approach to validate the final outcomes derived from GIS overlay and PCA. Thus, various remote sensing and GIS data were first explored in order to fully understand the concept of UEQ. The urban, environmental parameters and socio-economic indicators were normalized in this research in order to evaluate the significance of each parameter. GIS overlay and PCA (pixel-based and object-based) were introduced to integrate the urban, environmental parameters and socio-economic indicators with a case study in Toronto. Socio-economic indicators, including family income, degree of education and land value, were used as a reference to validate the outcomes derived from the two integration methods.

3.3 Datasets

In this research, the city of Toronto, Ontario, Canada, was intentionally selected due to the data availability and the drivers of the population growth within the city during the past decade. Figure 3.1 shows Toronto, which is the capital of the Province of Ontario and the largest city in Canada with a total population of 2,615,060 (Martel, 2012). The datasets being used in this study include three major categories: (1) Landsat TM satellite images; (2) GIS data layers; and (3) socio-economic data. All of the data were collected in the years 2010 and 2011, since GIS data and socio-economic are not consistently available after the year 2011. A Landsat TM image was downloaded from the USGS Earth Explorer (United States Geological Survey, 2014). The spatial resolution of the Landsat images is 30 m for the multi-spectral bands and 120 m for the thermal band. However, the thermal band was resampled to a 30-m resolution from the source of the data predominantly to align it with the multi-spectral bands (Kjaersgaard and Allen, 2009).

The image was acquired during the summer season (July) in order to avoid the appearance of clouds and snow cover. On the other hand, a total of 14 GIS data layers were acquired from (Scholars GeoPortal, 2014) for Toronto during the same period of time. The GIS layer data including land use, population density, building density, vegetation and parks, public transportation, historical areas, Central Business District (CBD), sports areas, religious and cultural zonse, shopping centres, education institutions, entertainment zones, crime rate and health condition. These layers were first imported into the ArcGIS platform (*ArcGIS*; Esri; Redlands, CA, USA) for further analysis. Similar to the remote sensing data, all of the data were projected to the Universal Transverse Mercator (UTM) 17 N coordinate system. Those social-economic indicators were derived based on the use of Toronto census data that were obtained from the City of Toronto census bureau at the census tract level. The City of Toronto census bureau archives hundreds of information related to socio-economic conditions. In this research, the socio-economic indicators included education (university certificate, diploma or degree), family income and land values. Table 4.1 summarizes the data sources being used in this study.

Table 3.1: The data sources for City of Toronto.

City	Landsat TM	GIS Data	Census Data
Toronto	Path/Row = 18/30 Sensor = Landsat TM Date = 23 June 2011 Remote sensing data: ∘ LST ∘ NDVI ∘ NDWI ∘ NDBI and Built-up Area	∘ Land Use ∘ Population Density ∘ Building Density ∘ Vegetation and Parks ∘ Public Transportation ∘ Historical Areas ∘ Central Business Districts ∘ Sports Areas ∘ Religious and Cultural Zones ∘ Shopping Centres ∘ Education Institutions ∘ Entertainment Zones ∘ Crime Rate ∘ Health Condition ∘ Areas Close to Water Bodies	Socio-economic data are provided by the City of Toronto census bureau. Socio-economic data used in the research: ∘ Education ∘ Family Income ∘ Land Values

LST, Land Surface Temperature; NDVI, Normalized Difference Vegetation Index;
NDWI, Normalized Difference Water Index; NDBI, Normalized Difference Built-up Index

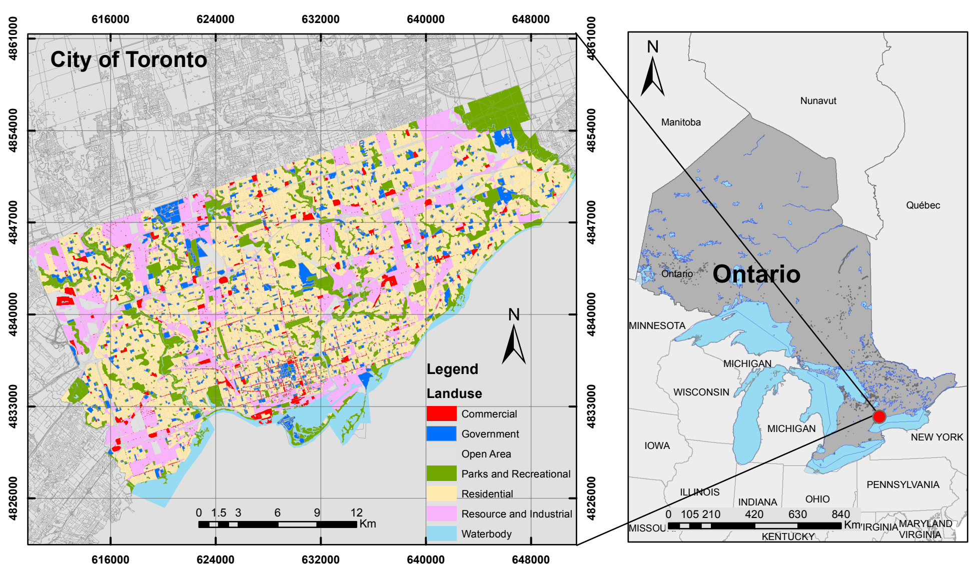


Figure 3.1: City of Toronto (the study area).

3.4 Methodology

Figure 3.2 shows the overall workflow implemented in this research. The Landsat image was clipped to the study area to speed up the data processing. The Atmospheric Correction model (ATCOR2) developed by (Richter, 1998) was utilized to preform radiometric calibration and remove the effects that change the spectral characteristics of the land features (Paolini *et al.*, 2006). To implement the ATCOR2 model, weather information (e.g., air temperature, visibility, etc.) was obtained from historical records at the nearest weather station at Lester B. Pearson International Airport. The calibration parameters for Landsat TM sensor (biases and gains) were also incorporated into the atmospheric correction.

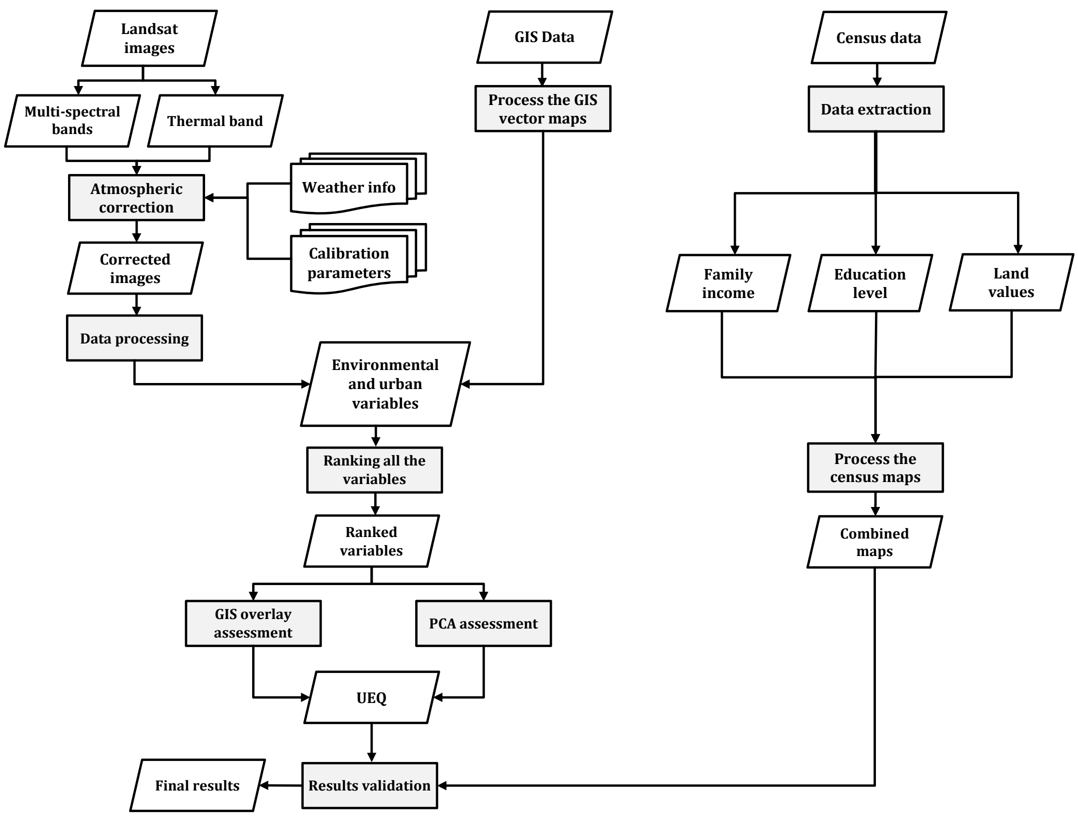


Figure 3.2: The overall workflow for investigating of GIS overlay and PCA techniques.

After conducting the atmospheric correction, those bio-physical parameters, including NDVI, NDWI, built-up index and LST, were derived from the Landsat images. Urban, environmental parameters and socio-economic indicators were all extracted from the remote sensing, GIS and census data to combine all of the parameters together in the subsequent process. GIS overlay and PCA (pixel-based and object-based approach) were implemented, respectively, to integrate all of the urban, environmental parameters and socio-economic indicators. Socio-economic indicators obtained from the City of Toronto census bureau, including family income, higher education level and land values, were used as a reference to assess the outcomes from GIS overlay and PCA. The validation was based on two criteria, including precision and accuracy (refer to Section 3.4.5). The final stage of the work is to assign the optimal integrated method to determine the best UEQ location in Toronto.

3.4.1 Environmental Parameters

Land Surface Temperature (LST)

LST is an essential parameter in a variety of disciplines used to study the urban climate (Norman *et al.*, 1995; Czajkowski *et al.*, 2004), UEQ (Nichol and Wong, 2009), urban heat island effect (Weng *et al.*, 2004), urban expansion (Huang *et al.*, 2015) and urban waste management (Yan *et al.*, 2014). LST is the result of a land-surface process that combines the analysis of all surface-atmosphere interactions and energy fluxes between the atmosphere and the ground. Mapping the LST from thermal remote sensing sensors can be useful for large-scale environmental and urban studies. Landsat TM and ETM+ data were substantially used in many urban environmental quality studies to derive the LST (Nichol and Lee, 2005; Nichol and Wong, 2006; Nichol *et al.*, 2006). Landsat TM and ETM+ both have: (1) an archive of images that was released free to the public by the (United States Geological Survey, 2014) in 2008 and (2) a short repeat cycle (16 days), which produces a voluminous data archive for multi-temporal studies. Numerous researchers discussed the use of LST and the challenges to retrieve the LST using known and unknown Land Surface Emissivity (LSE) (Li *et al.*, 2013; Sobrino *et al.*, 2004). In this research, the authors utilized PCI Geomatica (*Geomatica*, version 10.1; PCI Geomatics, Markham, ON, Canada, 2007). to derive the LST from the Landsat images. The adopted method to derive the LST in this research takes into consideration the atmospheric correction of the thermal band of the image. The computation of LST mainly involves three steps. The first step is to convert the pixel value of the thermal band into radiance using the following Equation (3.1):

$$L_{sat} = \Big[\frac{L_{max} - L_{min}}{(Q_{cal.max} - Q_{cal.min})}\Big](Q_{cal} - Q_{cal.min}) + L_{min} \tag{3.1}$$

where L_{sat} is the spectral radiance; L_{max} is the spectral radiance that is scaled to $Q_{cal.max}$; L_{min} is the spectral radiance to $Q_{cal.min}$; Q_{cal} is the quantized calibrated pixel value in a digital number; and $Q_{cal.max}$ is the maximum quantized calibrated pixel value corresponding to L_{max}. For Landsat TM Band 6, the values for L_{max}, L_{min} and $Q_{cal.max}$ are 15.3032 $Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$, 1.2378 $Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ and 255, respectively.

The second step is to compute the emissivity value. Many factors, including water content, chemical composition, structure and roughness, are able to affect the emissivity of a surface (Snyder *et al.*, 1998). Scholars emphasized that the surface temperature calculation mainly relies on an assumption of the emissivity value (Richter and Schläpfer, 2005). Some researchers assumed the emissivity value as a constant value (0.95) Coll *et al.* (2010). In contrast, other researchers epitomized that a constant emissivity value can be considered as an option and assigned three classes for the emissivity values, where the vegetation has $\epsilon = 0.97$, soil $\epsilon = 0.96$ and others $\epsilon = 0.98$ as a rule of thumb (Richter and Schläpfer, 2005). However, if the emissivity value is unknown, the following Equation (3.2) can be used to calculate the emissivity value (Giannini *et al.*, 2015):

$$\epsilon = a + b \times ln(NDVI) \tag{3.2}$$

where a and b are obtained by a regression analysis based on a large dataset (Moran *et al.*, 1992). NDVI is the Normalized Difference Vegetation Index, which can be calculated from the values of the visible and near-infrared bands of the multi-spectral bands, as shown in Section 3.4.1.

The third step is to conduct the atmospheric correction for the thermal band using the following Equation (3.3). As mentioned in Section 3.4, weather information (e.g., air temperature, visibility, etc.) and date and time, latitude and longitude are also needed to implement atmospheric correction. The equation for the atmospheric correction can be written as (Barsi *et al.*, 2005):

$$L_C = \frac{L_{sat} - L_{up}}{\epsilon \times \tau} - \frac{1 - \epsilon}{\epsilon} \times L_d \tag{3.3}$$

where L_C is the atmospherically-corrected radiance, L_{sat} is the spectral radiance ($\mathrm{Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}}$), L_{up} and L_d are the upwelling and downwelling radiances ($\mathrm{Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}}$) and ϵ and τ are the emissivity and transmittance, respectively.

The fourth step is to convert the calibrated radiance into the at-sensor brightness temperature using the following Equation (3.4):

$$L_{BBT} = \left[\frac{K_2}{ln(\frac{K_1}{L_C} + 1)} \right] \tag{3.4}$$

where T_{BBT} is the blackbody temperature in Kelvin (K), K_1 is the calibration Constant 1 in $\mathrm{Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}}$ and K_2 is the calibration Constant 2 in Kelvin (K). For Landsat TM, K_1 and K_2 are 607.76 $\mathrm{Wm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}}$ and 1260.56 K, respectively Chander *et al.* (2009).

The fifth step is to convert temperature from Kelvin into temperature in Celsius using the following Equation (3.5):

$$^{\circ}C = \left[L_{BBT} - 273.15 \right] \tag{3.5}$$

The computed (°C) is regarded as the LST derived from the Landsat image.

Normalized Difference Vegetation Index (NDVI)

Prior to the existence of satellite remote sensing, urban vegetation was usually monitored and mapped by combining colour infrared aerial images and fieldwork. This method seems to be a unique option to measure the urban vegetation (Nowak *et al.*, 1996). With the availability of multi-source multi-spectral satellite images, Fung and Siu (2001) used Landsat and SPOT (Satellite Pour lObservation de la Terre; Satellite for the Observation of Earth; Spot Image, Toulouse, France) images to quantify urban vegetation as a parameter for UEQ studies. Many researchers used Landsat images to extract NDVI (Lo *et al.*, 1997; Nichol and Lee, 2005; Nichol and Wong, 2009). NDVI is a ratio that presents the changes in the vegetation over time, and it has been applied to various applications, such as vegetation cover, biomass and Leaf Area Index (LAI) (Curran and Steven, 1983; Lawrence and Ripple, 1998). Most of the urban environmental studies showed that NDVI is one of the most important parameters that can be used to assess UEQ, where the higher values represent the positive impact on the city (Nichol and Lee, 2005; Nichol and Wong, 2009). The NDVI (ranging from −1 to 1) refers to an index that is able to monitor the vegetation activity and its annual changes, which can be calculated using Equation (3.6) (Zha *et al.*, 2003):

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red} \tag{3.6}$$

where NIR is the near infrared Band 4 in the Landsat TM image and Red is the red Band 3 in the Landsat TM image.

Normalized Difference Water Index (NDWI)

NDWI is another remote sensing-derived biophysical parameter that represents the surface moisture in vegetation cover, as well as water bodies. Hardisky *et al.* (1983) found that NDWI is able to track changes in vegetation biomass and water stress more than NDVI. NDWI can also be used to measure and assess the turbidity of water bodies from remote sensing data (McFeeters, 1996), and therefore, Liang and Weng (2011) used NDWI as a parameter to assess the UEQ where the higher NDWI represents the higher urban quality (i.e., close to lake shore). The NDWI (ranging from −1 to 1) can be are calculated using Equation (3.7) (Jensen, 2005):

$$\text{NDWI} = \frac{Green - NIR}{Green + NIR} \tag{3.7}$$

where NIR is the near infrared Band 4 in the Landsat TM image and $Green$ is the green Band 2 in the Landsat TM image.

Normalized Difference Built-Up Index (NDBI) and Built-Up Index

NDBI is another ratio that represents the spatial distribution of the urban and suburban areas. NDBI has been used in many urban planning applications. Zha *et al.* (2003) used the combination of NDBI and NDVI to identify and monitor the areas in the city of Nanjing. Chen *et al.* (2006) shows that land cover types can be represented by utilizing NDVI, NDWI and NDBI. Moreover, Faisal *et al.* (2016) and Faisal and Shaker (2014) show that the built-up index derived from NDBI and NDVI could represent industrial areas within the city. Therefore, in UEQ studies, the higher NDBI/built-up values may be deemed to have a negative impact on the city. To derive the built-up area, first, the NDBI values (ranging from -1 to 1) are calculated using Equation (3.8) (Zha *et al.*, 2003):

$$\text{NDBI} = \frac{MIR - NIR}{MID + NIR} \tag{3.8}$$

where MIR is the mid-infrared Band 5 of the Landsat TM image and NIR is the near infrared Band 4 of the Landsat TM image. The NDBI values refer to an index that represents the urban regions and its annual changes. Finally, the built-up values (ranging from -1 to 1) are defined by subtracting the NDBI layer from the NDVI layer using the following Equation (3.9) of Zha *et al.* (2003):

$$\text{Built-up area} = NDBI - NDVI \tag{3.9}$$

3.4.2 Urban Planning Parameters

Land Use and Land Cover

The expansion of population can affect the urban environment and urban planning around the world. Therefore, monitoring land use and land cover should be conducted to avoid potential problems for sustainable urban and environmental planning. Monitoring land use and land cover helps planners and decision makers to build better urban environmental cities in the near future and assess the quality of the urban cities. Various studies recommended building urban green cities rather than a dense high rise urban environment. Urban green cities increase the value of UEQ within the city (Irvine *et al.*, 2009; Landorf *et al.*, 2008; Din Özdemir, 2007). Medium to fine-scale land cover and land use maps can be derived from remote sensing satellite images (Hansen and Loveland, 2012) or, recently, airborne LiDAR data (Yan *et al.*, 2015). However, the accuracy of land cover and land use can change from one satellite to another due to the variation of the spatial resolutions of the satellites. In order to assess the urban quality of living, physical environmental parameters should be obtained. Physical environmental parameters, such as roads, cropland and pasture, water, commercial and industrial, high density residential, medium density residential, low density residential, forest and grass, are critical and essential parameters to assess the urban quality of life. The physical environmental parameters can be used also to extract some of the socio-economic indicators, such as population and social conditions (Liang and Weng, 2011).

Urban Density

Around the world, residential areas can be affected by the increase of population and migration movement. Building density is one of the most important parameters that contributes to the urban heat island effect and urban quality assessment (Mira *et al.*, 2005). Building and population density can have a negative influence on the UEQ and transportation system in the developing cities. That is mainly because a dense high rise urban environment typically increases LST, noise pollution together with a high demand of vehicle use (Kahn, 2007). However, most public services, public transportation and jobs are located within walking distance from high density areas. Remote sensing technique can aid in determining the density values by extracting the urban areas from the image (Rahman *et al.*, 2011; Nichol and Wong, 2009). The extracted urban areas can be divided by the total areas, so as to calculate the building density, as shown in Equation (3.10). On the other hand, the population density can be calculated by dividing the number of people over the urban area as shown in Equation (3.11):

$$\text{Building density} = \frac{\text{Urban areas}}{\text{Total areas}} \tag{3.10}$$

$$\text{Population density} = \frac{\text{Number of people}}{\text{Urban areas}} \tag{3.11}$$

Public Transportation

The acceleration of population growth may increase car ownership, which may increase the amount of carbon dioxide emission and subsequently affect the accessibility to roads, especially in the developing countries (Newman and Kenworthy, 1999). Transportation is the main sector that works in shaping and connecting the cities. Public transportation provides a faster, safer and easier way to travel around the city. Public transportation can help the city through connecting the sub-centres around the railway stations and building a linear development along the route of the public transit line (Newman and Kenworthy, 1999). It was found that most of the automobile-dependent cities lose the traditional community support processes (Newman and Kenworthy, 1999). Therefore, public transportation is one of the major parameters for the UEQ.

Open Spaces and Entertainment Zones

Many studies in UEQ justified that open spaces and open green areas are significant factors contributing to high environmental quality areas (Rahman *et al.*, 2011; Nichol and Wong, 2009). That is mainly because open spaces and parks offer a healthy and comfortable environment by cooling down the LST and reducing the air pollution especially in high density areas. Entertainment areas are mainly located in the public parks, plazas and open space areas for some occasions, such as Christmas and New Year. Famous open spaces, such as Times Square in the city of New York, Dundas square and Nathan Phillips Square in Toronto, are so invigorating with a big amount of visitors all over the year, mainly because they are located within the core of high density areas and thus provide a vibrant atmosphere. Such a phenomenon supports the argument that high density areas are more preferable than low density areas.

Historical Areas and Central Business Districts (CBD)

The design of historical cities around the world is mainly based on walking distance. Those historical cities are usually featured by high density, mixed land use and shaded streets in central forms, such as Jerusalem, Damascus, Athens and Istanbul. The average walking distance toward the historical cities is designed to be 5 km apart in order to be close to other facilitates in the city. A few cities still currently retain the historical buildings and walking characteristics, such as Society Hill in Philadelphia, the North End in Boston and the Rocks in Sydney (Newman and Kenworthy, 1999). That is mainly because historical areas retrieve the worth of past energy and provide a visual and physical conservation of cultural identity (Leask and Fyall, 2006). Currently, modern cities have more of a tendency to rebuild and preserve historical areas, such as Arabella Park in Munich, to attract tourists and provide a vibrant atmosphere for the city (Newman and Kenworthy, 1999). Historical neighbourhoods, which are always located in the city centre, have higher positive influence on UEQ, where the historical neighbourhoods and CBD are the most attractive regions in the city.

Crime Rate

Personal security is one of the most important factors for society regardless of where we live. Crime can be the reason for physical pain, anxiety and the loss of lives and property (Initiative, 2011). Anand *et al.* (2008) illustrated that the biggest influence of crime is the feeling of vulnerability in people's lives, and thus, the crime rate is negatively related to UEQ. It was reported that people move to live in more suburban and low density areas for the desire for new and better public schools and a low crime rate. However, in some cases, the low cost of housing may cause a demand for more housing per person, which may form new clusters for new urban crime (Cullen and Levitt, 1999). Increasing the physical distance between the poor and the rich is not always the best way to reduce urban crime, particularly in the city centre. Instead, it is preferable to increase the community services and the quality of life in those areas to make them more vibrant and reduce the crime rate (Kahn, 2007). The crime rate can be calculated by dividing the number of crimes over the total population, as shown in Equation (3.12):

$$\text{Crime rate} = \frac{\text{Number of crime}}{\text{Total population}} \tag{3.12}$$

3.4.3 Socio-Economic indicators

Education and Income

Education and income are two related factors among relevant socio-economic indicators. Research shows that wealthier urbanites tend to invest more in high quality properties and services. That is mainly because they have higher income and receive higher education, which gives them the tools to access and process more data about the high quality areas. In addition, people with high income and high education have the ability to invest in higher quality areas, compared to people with less education and less income (Becker and Mulligan, 1997). Moreover, Kahn (2007) pointed out that people with higher education and income are more interested in supporting UEQ-related issues. Wealthier and educated urbanites also tend to participate in politics and the community in order to enhance the quality of living in their living areas. Based on the above argument, the areas that have more highly-educated and wealthier urbanites are considered to have higher UEQ areas. Therefore, these areas are used as the first category of reference for our study.

Land Values

Knowing the parameters that influence the UEQ is an important advantage to design and assess the future urban development. UEQ is assessed by using various urban and environmental parameters. Reginster and Goffette-Nagot (2005) conducted a study in two Belgian cities to investigate the relationship between the UEQ with respect to the residential location. It was revealed that UEQ may affect positively the land rent location and income in the city. Other research discussed the relationship between the real estate evaluation model and the environmental parameters in the city of Geneva, Switzerland (Din *et al.*, 2001). It was found that urban and environmental parameters have an influence on the price within the city of Geneva. Topcu and Kubat (2009) also examined the relationship between urban and spatial factors that might influence the urban land values in the city of Istanbul. It was found that the distance from the sea, the distances from the central business district, universities and sanitary facilities, as well as the variable of the colour of building facades all have a predominant impact on the residential land values. As a result, our experiment assigned the land values as the second category of reference for this research.

3.4.4 Ranking the Parameters

Since the aforementioned parameters are extracted from different data sources, they may have different scale levels and cannot be combined to a specific unit. Therefore, all of the obtained data (parameters), including raster, census and GIS data, were first transformed into one scale (sub-neighbour), as shown in Figure 3.3. Then, all of the parameters were ranked from 1 to 10 to normalize the observation value for each parameter.

To normalize the parameters and represent the significant level of each polygon in the parameter, the Z-score method was performed for all parameters. The Z-score model is a statistical measurement that is able to standardize a wide range of data to represent the significant changes across the data (Cheadle *et al.*, 2003). Equation (3.13) shows the first step to normalize the parameters using the Z-score:

$$Z_i = \left[\frac{x_i - \mu}{\sigma}\right] \tag{3.13}$$

where x is the observation values (polygons) (refer to the GIS polygons of the parameters as shown in Figure 3.4), i is the parameter, μ is the mean value of the parameter and σ is the standard deviation of the parameter.

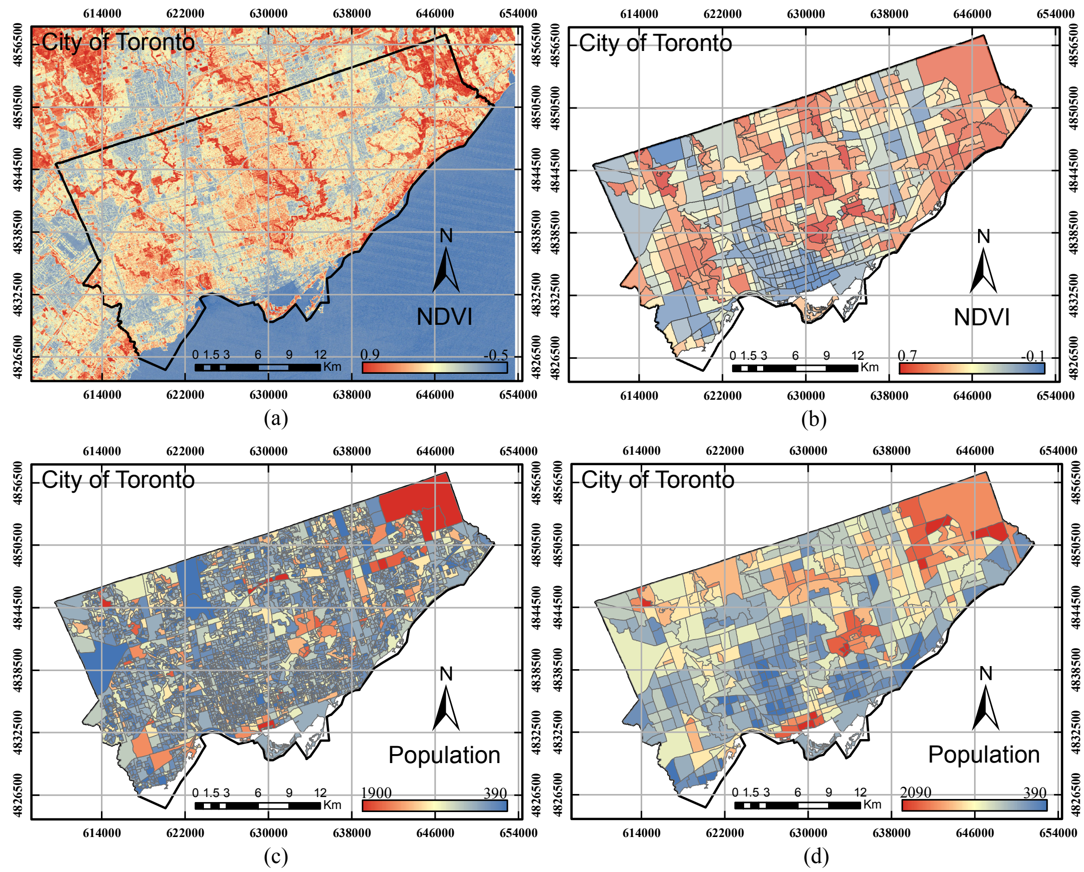


Figure 3.3: (**a**) NDVI image derived from Landsat image (raster data); (**b**) NDVI map after transformation (vector data); (**c**) population layer at the census tract level; (**d**) population layer after transformation to sub-neighbour level.

The second step is to use linear interpolation to rank the parameters from 1 to 10 as shown in Figure 3.5. The polygon within the parameter that has a high Z-score number will represent high values, for example 10. The polygon that has a low Z-score will result in a value of 1. The following Equation (3.14) shows how linear interpolation was calculated:

$$Rank = \left[\frac{(Obs - Obs_{max})(Rank_{min} - Rank_{max})}{(Obs_{min} - Obs_{max})}\right] + Rank_{max} \tag{3.14}$$

where Obs is the current observation value, Obs_{max} is the maximum observation value,

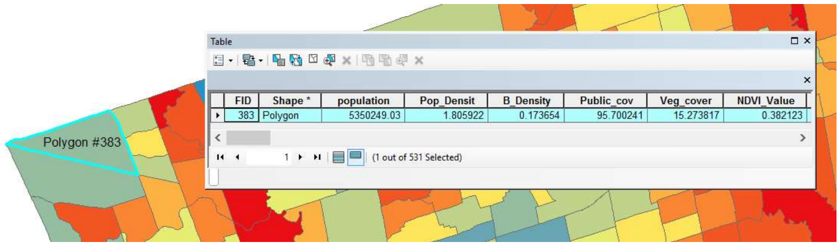


Figure 3.4: The GIS polygons of the parameters.

Obs_{min} is the minimum observation value, $Rank_{max}$ is the maximum ranking value, $Rank$ is the determined ranking value and $Rank_{min}$ is the minimum ranking value.

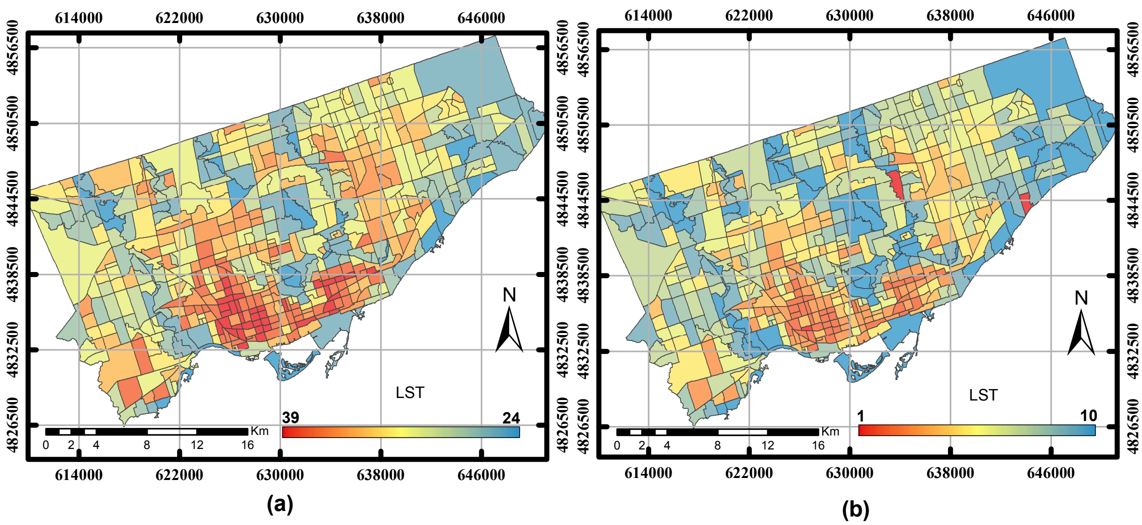


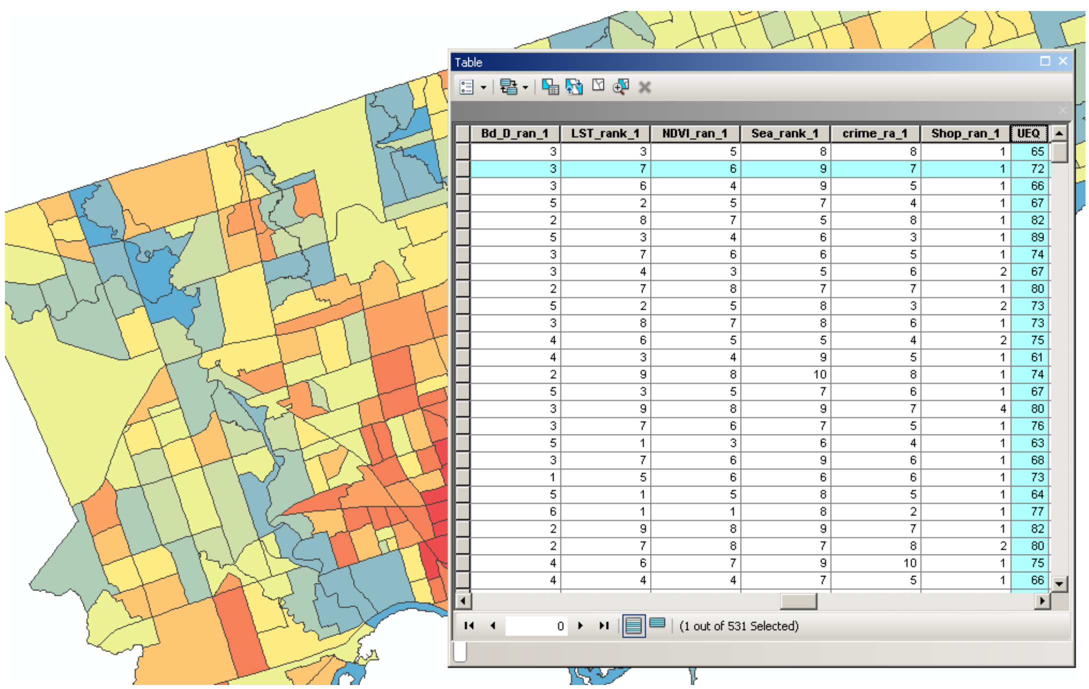
Figure 3.5: (**a**) The LST layer in degrees Celsius before ranking the parameter; (**b**) the ranking of LST after the normalization.

3.4.5 Data Integration of Multiple Environmental and Urban Parameters

Integration techniques can be used to combine remote sensing and GIS data and have been applied for urban modelling and analysis (Weng, 2002). Previous studies demonstrated two integration techniques, namely PCA and GIS overlay, which are able to combine any type of parameter. In this research, three approaches were demonstrated to integrate the above-mentioned environmental and urban parameters.

Geographic Information System (GIS) Overlay

GIS overlay is a multi-criteria application that uses data layers for specific environmental thresholds. Remote sensing data are presented as digital data in raster format. However, census data are presented in GIS vector format. Remote sensing data can thus be integrated with socio-economic data by converting remote sensing data from raster to vector data (Li and Weng, 2007). In this research, the GIS overlay integration method was used to combine the urban and environmental parameters in order to serve for the UEQ assessment. All of the parameters were converted from raster to vector data in order to be presented as attribute data, as shown in Figure 3.3 in Section 3.4.4. While each parameter has a range of values ranked from 1 to 10, the sum of the data layers can thus present the result of UEQ values. Ranking the parameters was mainly based on the observation values; where the highest value is assigned 10 and the lowest value is assigned 1. However, some parameters, including crime rate, industrial areas and LST, are inversely presented (e.g., the highest crime rate or LST value will be assigned 1, and the lowest crime rate or LST value will be assigned 10), as shown in Figure 3.5. Then, all of the ranks are summed up to compute the UEQ as shown in Figure 3.6.



Bd_D_ran_1	LST_rank_1	NDVI_ran_1	Sea_rank_1	crime_ra_1	Shop_ran_1	UEQ
3	3	5	8	8	1	65
3	7	6	9	7	1	72
3	6	4	9	5	1	66
5	2	5	7	4	1	67
2	8	7	5	8	1	82
5	3	4	6	3	1	89
3	7	6	6	5	1	74
3	4	3	5	6	2	67
2	7	8	7	7	1	80
5	2	5	8	3	2	73
3	8	7	8	6	1	73
4	6	5	5	4	2	75
4	3	4	9	5	1	61
2	9	8	10	8	1	74
5	3	5	7	6	1	67
3	9	8	9	7	4	80
3	7	6	7	5	1	76
5	1	3	6	4	1	63
3	7	6	9	6	1	68
1	5	6	6	6	1	73
5	1	5	8	5	1	64
6	1	1	8	2	1	77
2	9	8	9	7	1	82
2	7	8	7	8	2	80
4	6	7	9	10	1	75
4	4	4	7	5	1	66

Figure 3.6: The summed up ranks for all of the parameters.

Principal Component Analysis (PCA)

PCA is an analysis technique that compresses the high dimension of data into a lower dimension of data that has most of the variance of the data (Jensen, 2005). PCA is commonly used in many remote sensing applications. The covariance matrix of standard PCA may not be the best option for data that have different measurement units. The correlation matrix can be used instead of the covariance matrix to standardize each parameter to the variance unit or zero mean. In this research, pixel-based and object-based methods were used to assess the UEQ in Toronto. In pixel-based approach, all of the parameters were converted to raster format to extract pixel values for each parameters. Then, the pixel values were used in the PCA model to compute the components that have most of the variance of the data. In object-based PCA, the covariance matrix or correlation matrix mainly is derived from the observation values of the GIS polygons. Then, the covariance matrix or correlation matrix will be used to compute the components in the PCA model to assess the UEQ.

Accuracy Assessment

Several researchers attempted to assess the accuracy of the UEQ results using different methods, including e-mail questionnaires, field-based questionnaires and factor analyses. Regardless of the considerable amount of e-mail questionnaires or field-based questionnaires, both methods require overheads for data collection. In addition, factor analysis used in previous work was preformed using the same parameters that have been incorporated to compute the UEQ, which make it unreliable and biased. Several researchers illustrated that education level, including university certificate or diploma, family income and land values, represents the UEQ in the economic and social aspects (Becker and Mulligan, 1997; Kahn, 2007; Reginster and Goffette-Nagot, 2005; Din *et al.*, 2001). Since there is a lack of ground truth to validate the results, we propose to use these socio-economic indicators for data validation and to assess the UEQ results. All of the observation data of the three socio-economic indicators were normalized to be in the same scale from 1 to 10. Then, the sum of the socio-economic indicators can thus present the result of reference, as shown in Table 3.2.

Table 3.2: The sum of the socio-economic indicators.

Polygon ID	**Income**	**Education**	**Land Value**	**Reference Layer**
1	8	5	7	20

In addition, the evaluation of the binary classifiers approach was used to assess the UEQ based on the following two performance measures through data interpretation: precision and accuracy.

Precision (P) is a measure that evaluates the probability that a positive outcome is correct using Equation (3.15):

$$P = \left[\frac{\mid TP \mid}{\mid TP \mid + \mid FP \mid}\right] \tag{3.15}$$

Accuracy (Acc) evaluates the effectiveness of the classifier by its percentage of correct predictions using Equation (3.16):

$$Acc = \left[\frac{\mid TN \mid + \mid TP \mid}{\mid FN \mid + \mid FP \mid + \mid TN \mid + \mid TP \mid}\right] \tag{3.16}$$

where TP refers to "True Positive", which means the polygon from the proposed method is located physically in the reference layer; TN refers to "True Negative", which represents the polygons that are not detected in the proposed method and reference layer; FP refers to "False Positive", which means that the polygon of the proposed method does not really exist in the reference layer; and FN refers to "False Negative", which means the reference polygons do not exist in the proposed method. With these three indicators, we assessed the UEQ layer from the results of each proposed method including GIS overlay, and PCA assessed the best method for our datasets.

3.5 Results and Analysis

3.5.1 GIS Overlay Analysis

Figure 3.7 shows the UEQ derived in Toronto using the GIS overlay. The distribution of UEQ in Toronto shows that the highest UEQ zones were found in the zones A, B, C and D in green colour, while the lowest UEQ zones are indicated as red colour in the city. The highest UEQ zones are the consequences of the summation of all of the positive parameters including (high vegetation areas, historical areas, areas supported by public transportation, etc.) that are located within Zones A to D. However, negative values of the parameters, including crime, industrial areas and high LST, are constantly located on the red zones within the city. In contrast, the highest values of UEQ areas were found in the high and moderate density areas, while the lowest values were found in the industrial and low density areas.

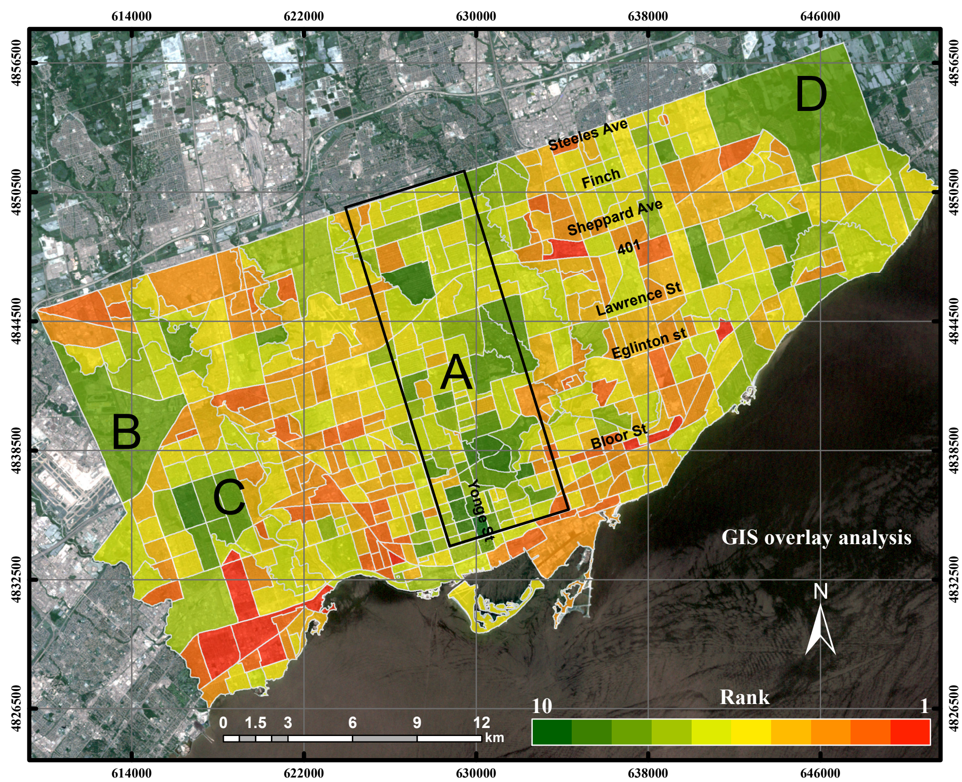


Figure 3.7: The UEQ derived using the GIS overlay method.

3.5.2 Principal Component Analysis

Pixel-Based PCA

In this section, an analysis was first conducted to investigate the relationship among all of the parameters. In pixel-based PCA, all of the parameters were converted from vector to raster in order to compute the spatial correlation among the parameters. Some parameters, including built-up areas, LST layer, industrial areas and crime rate regions, were reversed in order to avoid any negative values in the correlation matrix. Pearson's correlation coefficient was computed to investigate the dependence among all of the parameters, which is going to help in the subsequent PCA. Table 3.3 represents the correlation coefficient matrix among all of the parameters. The green vegetation parameter shows a strong positive relationship with NDVI (0.85), NDWI (0.85), reverse built-up areas (0.81) and reverse LST (0.90), as well as the areas close to water bodies (0.8). The green areas parameter also has a moderate correlation with the reverse industrial areas (0.69) and the reverse crime rate parameter (0.75). On the other hand, NDVI has a strong positive relationship with NDWI (0.98), reverse built-up areas parameter (0.96), reverse LST (0.91), green vegetation (0.85), the areas close to water bodies (0.80), reverse industrial areas (0.84) and the reverse crime rate parameter (0.80).

The reverse built-up areas parameter has a strong positive correlation with NDVI (0.96), NDWI (0.96), reverse LST (0.87), green vegetation (0.81) and reverse industrial areas (0.83). The areas that are close to water bodies (0.76) and the reverse crime rate parameter (0.80) both have a moderate correlation (0.76 and 0.77), respectively, with the reverse built-up areas parameter. The reverse crime rate parameter has a strong positive relationship with NDVI (0.80), NDWI (0.82), reverse LST (0.79) and the areas close to water bodies (0.82). On the other hand, the reverse crime rate also has a moderate correlation with reverse industrial areas (0.77), reverse built-up areas (0.77), green vegetation (0.75) and the public transportation parameter (0.70). Based on these observations, one can indicate that the high vegetation areas are usually located at low crime rate and low industrial areas within the city. The parameter of low crime rate is also influenced by the transportation within the city because of a high correlation observed between these two parameters. The areas that are covered by public transportation are usually crowded with people, which thus influences the crime rate within the city. These observations also indicate that the built-up areas have a high correlation with industrial areas, which could help to derive the industrial areas using remote sensing data. The high correlation between the parameters may cause redundancy and slow down the processing steps. Therefore, data reduction can help to improve the data processing and cost.

Table 3.3: The correlation coefficient matrix among all of the parameters derived using the pixel-based method.

	PD	**BD**	**PT**	**Veg**	**NDVI**	**NDWI**	**rBU**	**rLST**	**H**	**rInd**	**CBD**	**Sc**	**Ent**	**He**	**Rel**	**SP**	**Sea**	**rCR**	**SH**
PD	1.00	0.68	0.57	0.33	0.39	0.40	0.42	0.32	0.67	0.52	0.56	0.33	0.46	0.40	0.22	0.33	0.44	0.43	0.42
BD		1.00	0.62	0.33	0.40	0.42	0.60	0.36	0.48	0.52	0.40	0.43	0.32	0.44	0.45	0.33	0.59	0.60	0.41
PT			1.00	0.48	0.52	0.54	0.47	0.50	0.31	0.60	0.27	0.41	0.22	0.41	0.41	0.29	0.70	0.70	0.34
Veg				1.00	0.85	0.85	0.81	0.90	0.25	0.69	0.25	0.51	0.28	0.35	0.44	0.49	0.80	0.75	0.35
NDVI					1.00	0.98	0.96	0.91	0.21	0.84	0.16	0.52	0.18	0.33	0.33	0.43	0.80	0.80	0.22
NDWI						1.00	0.96	0.91	0.21	0.86	0.17	0.51	0.16	0.32	0.32	0.42	0.81	0.82	0.20
rBU							1.00	0.87	0.26	0.83	0.20	0.49	0.17	0.34	0.27	0.42	0.76	0.77	0.22
rLST								1.00	0.22	0.75	0.24	0.52	0.25	0.35	0.44	0.46	0.83	0.79	0.31
H									1.00	0.26	0.82	0.39	0.64	0.43	0.31	0.41	0.26	0.21	0.55
rInd										1.00	0.22	0.38	0.17	0.30	0.19	0.26	0.76	0.77	0.20
CBD											1.00	0.30	0.53	0.35	0.23	0.32	0.28	0.17	0.48
Sc												1.00	0.35	0.46	0.62	0.68	0.43	0.51	0.44
Ent													1.00	0.37	0.48	0.40	0.24	0.19	0.74
He														1.00	0.37	0.39	0.41	0.35	0.48
Rel															1.00	0.58	0.42	0.46	0.57
SP																1.00	0.38	0.41	0.43
Sea																	1.00	0.82	0.34
rCR																		1.00	0.33
SH																			1.00

PD, Population Density; BD, Building Density; PT, Public Transportation; Veg, Vegetation areas; rBU, reverse Built-Up areas; rLST, reverse LST; H, Historical areas; rInd, reverse Industrial areas; Sc, School areas; Ent, Entertainment areas; He, Health condition; Rel, Religion areas; SP, Sport areas; Sea, areas close to the Sea; rCR, reverse Crime Rate areas; SH, Shopping areas.

Four components were extracted from all of the parameters using the pixel-based PCA approach. Figure 3.8 shows the UEQ derived using the pixel-based PCA method. PC1 represents the largest percentage of the variance of the data, with 95% of the total variance. However, the combination of Components 2, 3 and 4 contains only 5% of the total variance. Due to the higher variance of Component 1, it represents most of the parameters, including crime rate, NDVI, NDWI, reverse LST, areas close to water bodies, reverse industrial areas, reverse built-up areas, green vegetation and public transportation parameter, as shown in Table 3.4. The low variance found in Components 2, 3 and 4 showed that the used pixel-based PCA relied only on the first components, as shown in Figure 3.9.

Table 3.4: The parameters vs. the components in the pixel-based PCA.

	Component 1	**Component 2**	**Component 3**	**Component 4**
Population Density	0.63	0.59	−0.35	−0.03
Building Density	0.31	0.46	0.16	−0.59
Public Transportation	0.90	0.01	−0.11	0.20
Vegetation areas	0.35	0.53	0.19	−0.60
NDVI	0.46	0.43	0.18	−0.23
NDWI	0.87	−0.19	−0.25	−0.22
Reverse Built-up areas	0.91	−0.22	0.20	0.04
Reverse Industrial	0.90	−0.31	0.08	−0.14
Reverse LST	0.93	−0.29	0.08	−0.04
Historical	0.93	−0.29	0.04	−0.04
CBD	0.54	0.42	−0.19	−0.48
School	0.73	0.44	−0.44	0.15
Entertainment	0.49	0.51	0.47	0.40
Health Condition	0.60	0.31	0.42	0.07
Religion	0.91	0.01	−0.10	0.09
Sport	0.40	0.56	0.36	−0.16
Sea	0.51	0.30	0.53	0.03
Reverse Crime rate	0.31	0.50	0.39	−0.34
Shopping	0.87	−0.17	0.25	0.05
Variance	95.00	2.53	2.36	0.11

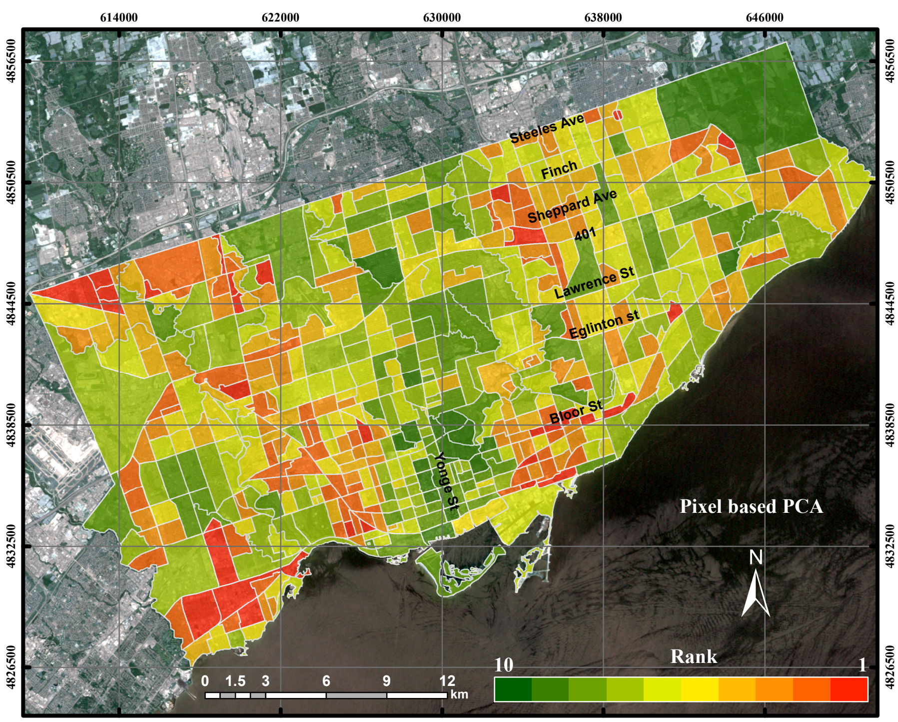


Figure 3.8: The UEQ derived using the first component of the pixel-based PCA method.

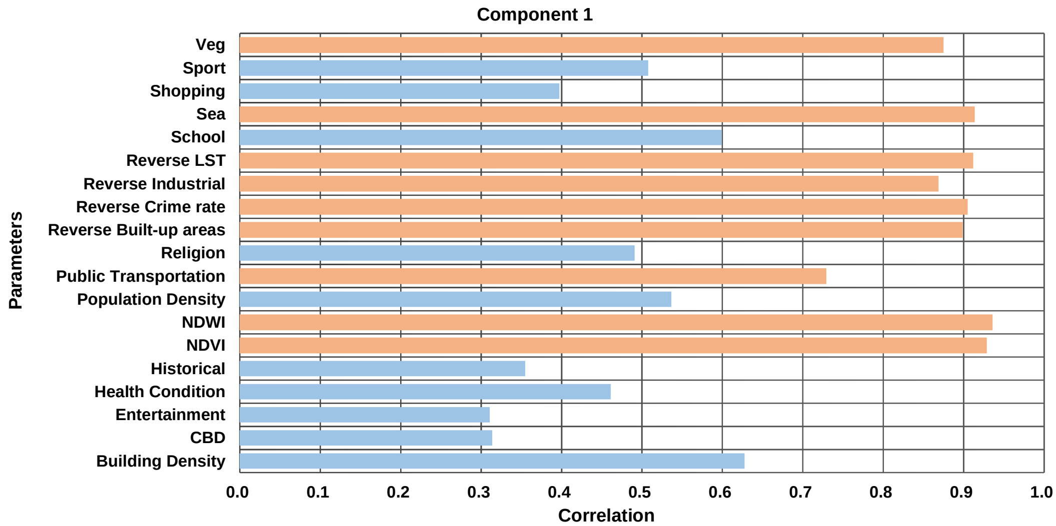


Figure 3.9: The UEQ parameters versus PCA Component 1.

Object-Based PCA

In the object-based approach, the polygons of each parameter were used in the PCA model to assess the UEQ. Table 3.5 represents the correlation coefficient matrix among all of the parameters. Population density has a moderate positive correlation coefficient with the historical areas parameter (0.66), where building density has a moderate negative correlation with green vegetation (-0.61), NDVI (-0.68), NDWI (-0.67) and a positive correlation with built-up areas (0.67) and LST (0.78). NDVI has a strong positive relationship with NDWI (0.88) and a moderate negative correlation with green vegetation (0.66). However, NDVI has a high negative correlation with the built-up areas parameter (-0.90) and LST (-0.80) and also has a moderate negative correlation with building density (-0.68). The built-up areas parameter has a strong positive correlation with building density (0.67) and LST (0.79). In addition, the built-up areas parameter has a negative correlation with NDVI (-0.90) and NDWI (-0.89). NDVI has a very high correlation with NDWI and a negative correlation with the built-up areas parameter and LST, as well as having a moderate negative correlation with building density, which indicates that high NDVI values represent low LST and low high building density areas with more green areas. As mentioned in the previous section, data reduction can improve the data processing and cost. Therefore, the object-based approach was used to reduce the size of the data.

Table 3.5: The correlation coefficient matrix among all of the parameters for the object-based method.

	PD	**BD**	**PT**	**Veg**	**NDVI**	**NDWI**	**BU**	**LST**	**H**	**Ind**	**CBD**	**Sc**	**Ent**	**He**	**Rel**	**SP**	**Sea**	**CR**	**SH**
PD	1.00	0.34	0.14	−0.14	−0.11	0.11	0.12	0.12	0.66	−0.04	0.08	−0.17	−0.02	0.03	−0.11	−0.04	−0.06	0.02	−0.04
BD		1.00	0.40	−0.61	−0.68	−0.67	0.67	0.78	0.44	0.07	0.39	−0.05	0.14	0.11	0.16	0.02	0.21	0.22	0.05
PT			1.00	−0.37	−0.37	−0.36	0.38	0.46	0.12	0.15	0.16	−0.09	−0.04	−0.01	0.05	−0.03	0.12	0.12	0.04
Veg				1.00	0.66	0.55	−0.56	−0.66	−0.11	−0.13	−0.09	−0.03	0.05	−0.03	−0.13	0.03	−0.30	−0.11	−0.02
NDVI					1.00	0.88	−0.90	−0.80	−0.30	−0.37	−0.37	0.02	−0.27	−0.10	−0.29	−0.09	−0.27	−0.35	−0.23
NDWI						1.00	−0.89	−0.77	−0.31	−0.39	0.37	−0.02	0.29	0.11	0.31	0.10	0.25	−0.35	0.26
BU							1.00	0.79	0.30	0.50	0.35	−0.01	0.27	0.10	0.31	0.09	0.27	0.35	0.24
LST								1.00	0.18	0.19	0.25	−0.02	0.05	0.05	0.14	0.00	0.31	0.19	0.06
H									1.00	−0.01	0.50	−0.05	0.43	0.24	0.09	0.16	−0.05	0.33	0.19
Ind										1.00	0.03	0.02	0.08	−0.01	0.31	0.05	0.06	0.12	0.14
CBD											1.00	−0.05	0.37	0.19	0.07	0.09	−0.07	0.38	0.16
Sc												1.00	0.04	0.12	0.25	0.05	0.21	0.00	0.03
Ent													1.00	0.30	0.26	0.39	0.00	0.38	0.49
He														1.00	0.30	0.49	- 0.03	0.21	0.38
Rel															1.00	0.44	0.11	0.15	0.41
SP																1.00	0.02	0.18	0.62
Sea																	1.00	0.01	0.03
CR																		1.00	0.27
SH																			1.00

PD, Population Density; BD, Building Density; PT, Public Transportation; Veg, Vegetation areas; BU, Built-Up areas; LST, LST; H, Historical areas; Ind, Industrial areas; Sc, School areas; Ent, Entertainment areas; He, Health condition; Rel, Religion areas; SP, Sport areas; Sea, areas close to the Sea; CR, Crime Rate areas; SH, Shopping areas.

In this study, five components were extracted in the object-based PCA approach, which have eigenvalues larger than one, as shown in Figure 3.10. The total variance of the five components is 75% of the overall variance of the data. Preliminary analysis revealed that Component 1 has 36% of the total variance of the dataset. Component 1 shows strong positive loadings with NDVI (0.88), NDWI (0.86), building density (0.80) and historical areas (0.86) and strong negative loadings with LST (−0.86) and built-up areas (−0.86). In addition, Component 1 is the best to represent the green areas within the city. Component 2 reveals about 16% of the dataset, which mainly represents industrial areas with a positive correlation of 0.63 and CBD with a positive correlation of 0.76. Component 2 can be used to represent more about the urban areas. Component 3 represents 9% of the dataset, which mainly represents only sports areas with a positive correlation of (0.81). Component 4 reveals 7% of the dataset, which mainly represents public transportation with a positive correlation of 0.70. Table 3.6 shows the overall map produced from Components 1 to 5, which represents 75% of the overall variance in the data.

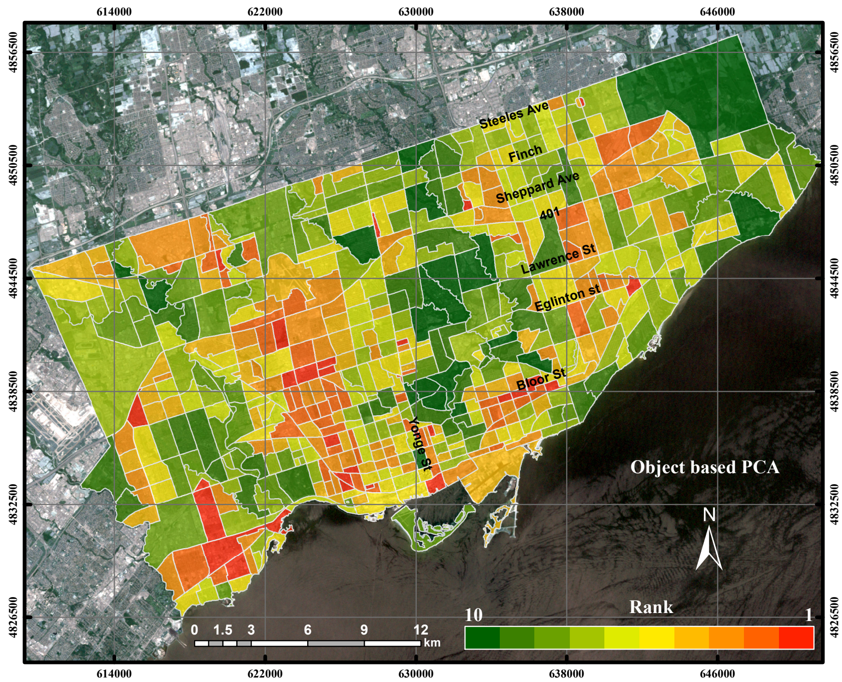


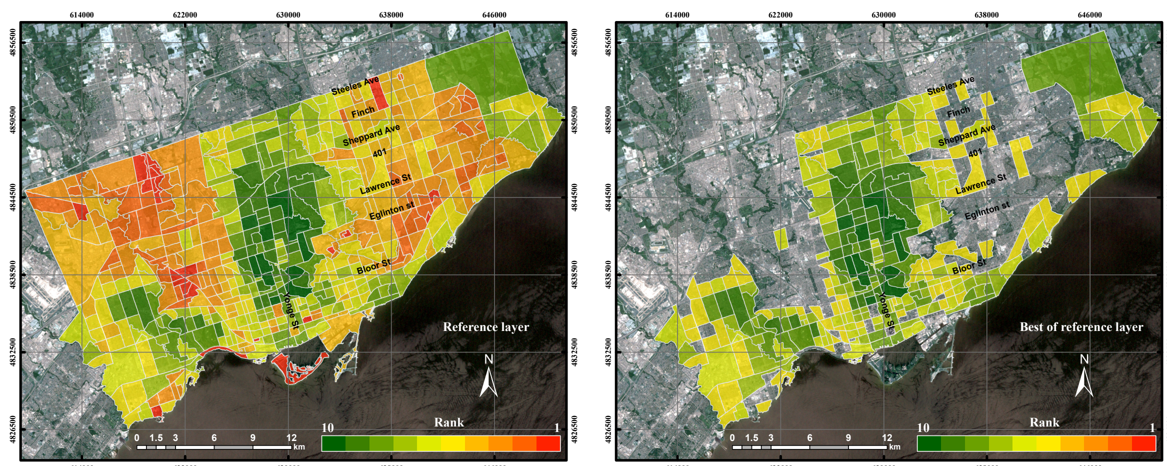
Figure 3.10: The UEQ derived using four components of the object-based PCA method.

Table 3.6: The parameters vs. the components in the object-based PCA.

	Component 1	**Component 2**	**Component 3**	**Component 4**	**Component 5**
Population Density	−0.41	0.04	−0.46	0.14	0.16
Building Density	0.80	−0.09	−0.08	0.10	0.13
Public Transportation	0.49	0.00	0.00	0.70	−0.24
Veg	0.69	0.42	0.17	−0.07	−0.11
NDVI	0.88	0.12	−0.06	0.26	−0.09
NDWI	0.86	−0.13	0.08	0.27	0.08
Built-up areas	−0.86	0.18	−0.07	0.27	−0.09
Industrial	−0.59	0.63	0.00	0.16	0.08
LST	−0.86	0.34	0.05	0.03	−0.11
Historical	0.86	0.35	0.05	0.29	−0.13
CBD	0.56	0.76	−0.02	0.15	0.06
School	0.04	−0.11	0.54	0.02	−0.29
Entertainment	−0.28	0.43	0.31	−0.07	−0.02
Health Condition	−0.14	0.21	0.17	−0.08	−0.01
Religion	−0.26	−0.07	0.48	−0.20	−0.01
Sport	0.02	0.04	0.81	0.22	0.40
Sea	−0.36	−0.42	0.29	0.34	−0.54
Crime rate	0.44	−0.35	−0.04	0.48	0.53
Shopping	−0.18	0.14	0.33	−0.11	0.07
Variance	35.83	15.97	8.97	7.24	6.83

3.5.3 UEQ Validation Results

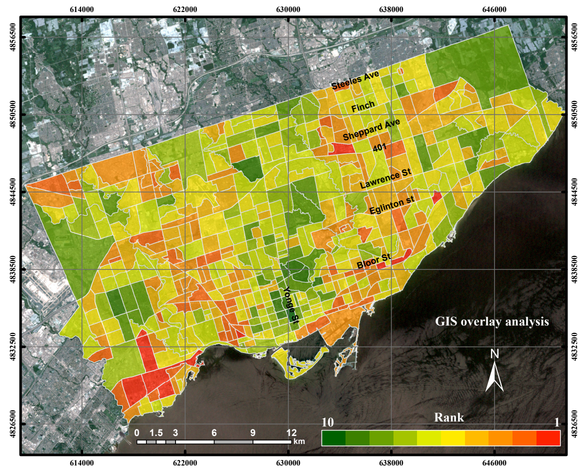
As mentioned in the previous section, three socioeconomic indicators were derived from census data. The combination of education level, family income and land values was used to validate the UEQ results. The evaluation of binary classifiers approach was used to evaluate the UEQ, as mentioned in Section 3.4.5. The results of GIS overlay and PCA (pixel-based and object-based) were validated using socioeconomic indicators as a reference for this study. Since we are looking to highlight the higher UEQ areas, the mean values were used as a threshold to derive the higher UEQ areas. Figure 3.11 shows the reference layer and the high value of the reference layer. The distribution of the reference layer revealed that the highest values are found in the city centre, the west portions of the city, while most of the low UEQ values are found in the east and down town of the city.



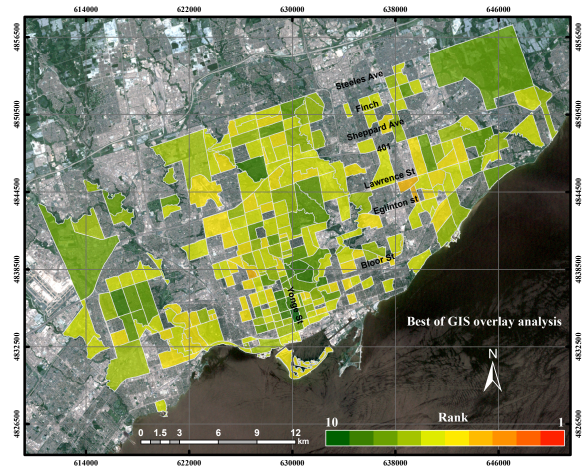
(**a**) (**b**)

Figure 3.11: The reference layer and the results of the reference layer higher than the mean. (**a**) The reference layer; (**b**) results of the reference layer higher than the mean.

Figure 3.12 shows the GIS overlay analysis and the higher values of GIS overlay. There exist a few areas having high UEQ values located in the north and east of the city. The precision and accuracy measured were found to be 71% and 65%, respectively, for the GIS overlay method. That is mainly because the GIS overlay method uses all of the parameters where some of the parameters may have a negative correlation with the reference layer, which may influence the overall result.



(**a**)

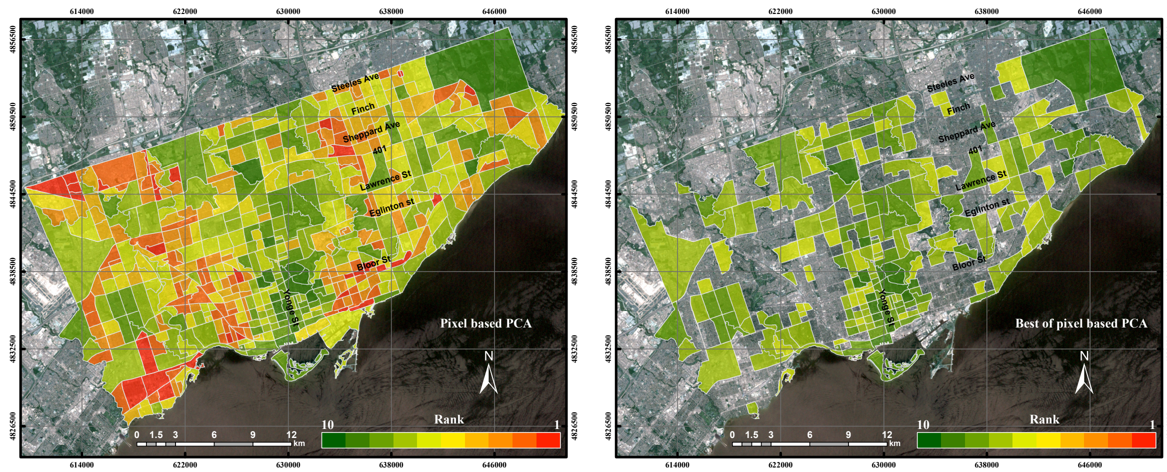


(**b**)

Figure 3.12: The UEQ derived using the GIS overlay method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean.

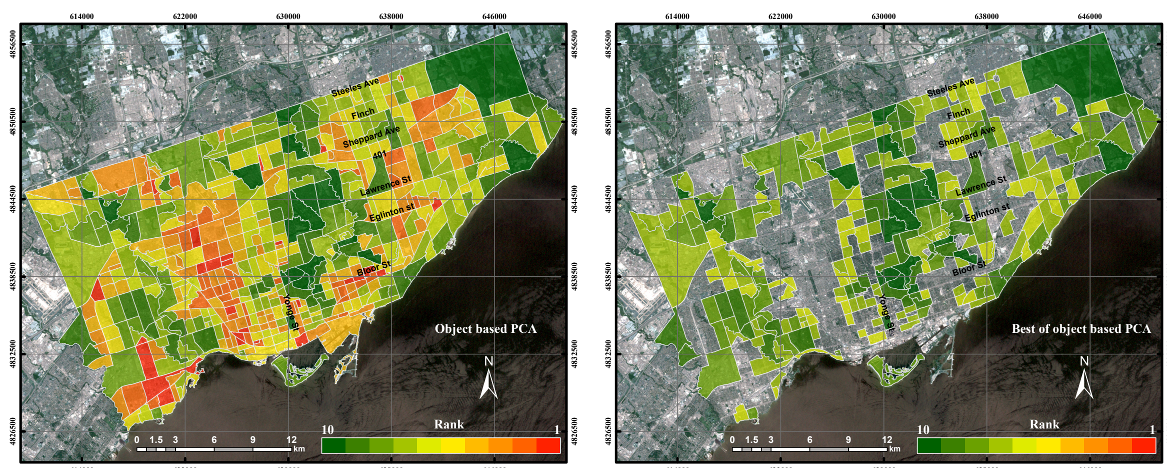
Figure 3.13 shows higher UEQ ranking derived using the pixel-based PCA method. The highest values of pixel-based PCA are mainly located in the centre, north, northwest and northeast portions of the city. Since the pixel-based PCA used 95% of the data, the result of the pixel-based PCA shows lower precision and accuracy with respect to GIS overlay. The precision and accuracy are reported to be 68% and 63%, respectively, for pixel-based PCA. Apparently, the pixel-based PCA reveals a lower precision and accuracy than GIS overlay, mainly because the pixel-based PCA considered only nine parameters to generate 95% of the data, and some of these parameters have low correlation with the reference layer.

Figure 3.14 shows the object-based PCA and the higher values of the object-based PCA. The result of the object-based PCA represents high UEQ values in the centre, north, northwest and northeast portions of the city. The overall result of object-based PCA reveals a slightly better precision and accuracy by 1% than the pixel-based PCA method. The main reason why the object-based PCA results were slightly better than the pixel-based PCA is mainly because the object-based PCA method considered five components in the analysis, which have more variation of the parameters. However, only one component was considered in the analysis in pixel-based PCA. One more reason could be because in pixel-based PCA, all of the vector data were converted to raster data. That step may cause a certain loss of spatial information, which may affect the overall results.



(**a**) (**b**)

Figure 3.13: The UEQ derived using the pixel-based method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean.



(**a**) (**b**)

Figure 3.14: The UEQ derived using the object-based method. (**a**) The derived UEQ; (**b**) UEQ zones higher than the mean.

The overall result of the object-based PCA method yielded a lower precision and accuracy by 1% than the GIS overlay method, as shown in Figure 3.15, and that is mainly because of the same reason for pixel-based PCA, which is the object-based PCA method used only 75% of the total variance. However, the GIS overlay method used all of the parameters.

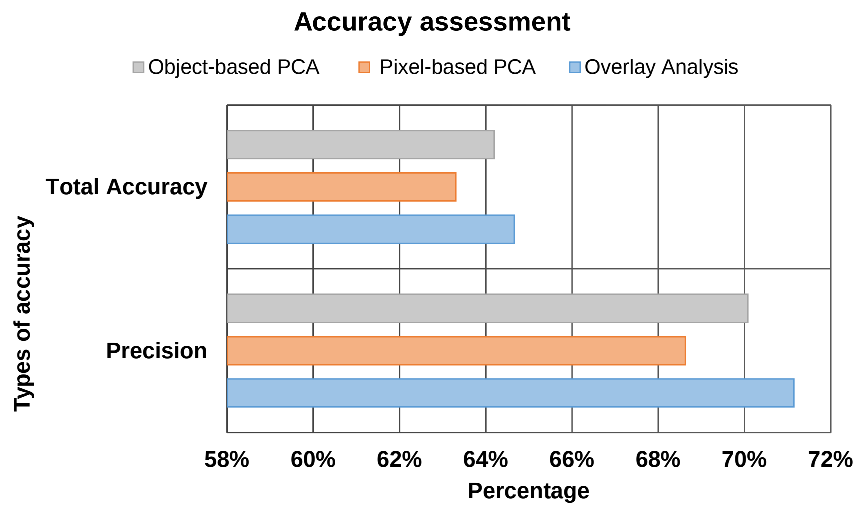


Figure 3.15: The UEQ validation.

3.6 Conclusions

In summary, this study aimed to utilize remote sensing and GIS techniques to assess UEQ with a case study in the city of Toronto, Ontario, Canada, through evaluating two methods: GIS overlay and PCA. One of the issues for the UEQ integration method is that remote sensing, GIS and census data are collected at different scales and in different formats, which may require data normalization before further analysis. In this study, The Z-score model was performed as a first step to normalize all of the parameters. Then, linear interpolation was implemented to rank all of the Z-score values from 1 to 10.

Integration techniques including GIS overlay and PCA (both pixel-based and object-based methods) were used to integrate the environmental, urban parameters and socio-economic indicators. GIS overlay is one of the effective tools for integrating different datasets from different data sources. GIS overlay offers an intelligent platform for creating a comprehensive database to evaluate the UEQ. Correlation analysis investigates the dependence found among urban, environmental parameters and socioeconomic indicators.

In our case study, it was found that green areas have a strong positive correlation with NDVI and NDWI. There was a negative relationship with the built-up areas parameter, LST, industrial areas, crime rate and building density. Alternatively, PCA provides an efficient method to reduce the data dimension and redundancy. Four components that have eigenvalues over one were derived from the 19 parameters that represented the urban and environmental aspects in the pixel-based PCA method. Five components that have eigenvalues over one were derived from the 19 parameters that represent the urban and environmental aspects in the object-based PCA method. The two methods (pixel-based and object-based) were tested due to the data availability. Other studies can only consider one method of PCA, since they do not have significant contrast in the results with respect to UEQ parameters.

One of the key concerns in UEQ research is to validate the final results derived from different socio-economic references. Despite that some of the existing UEQ studies utilized email or questionnaire surveys to collect the public's opinion for UEQ assessment, this study proposed to use three socio-economic indicators (university certificate or diploma, family income and land values) as a reference for result assessment. The results showed that the precision was 71% for the GIS overlay method, and the accuracy was measured as 65%. The precision level of the pixel-based PCA method yielded 68%, and the accuracy was reported to be 63%, respectively. The precision level of the object-based PCA was 70%, where the accuracy was reported to be 64%. In this study, GIS overlay represented better results than PCA (pixel-based and object-based) with respect to the UEQ results parameters, which may suggest that GIS overlay can be a better method in terms of the integration of multiple parameters.

Although the presented approach can be used by any federal authorities and municipalities in developing and developed countries, where there is a need to improve and design the new areas within the city, there are a few recommendations for similar future studies: (1) more up-to-date remote sensing and GIS data are required to consolidate the findings; (2) census socioeconomic data usually relate to administrative units and can be changed in a shorter period of time, which makes it difficult to be available worldwide; (3) integration among remote sensing, GIS and socioeconomic data needs conversion between data, such as from raster to vector or from vector to raster, a step that may cause a certain loss of spatial information. To conclude, remote sensing and GIS techniques can provide fruitful information to model UEQ. However, other urban and environmental parameters, as well as empirical models (such as different geographically-weighted approaches) should be considered in order to develop a more universal indicator to predict the UEQ. As a result, further research is under way to study different approaches to narrow down the variety of parameters, as well as developing a new technique to retrieve the UEQ in different cities located in Canada.

Chapter 4

Improving the Accuracy of UEQ Assessment Using Geographically-Weighted Regression Techniques

4.1 Abstract

Urban Environmental Quality (UEQ) can be treated as a generic indicator that objectively represents the physical and socio-economic condition of the urban and built environment. The value of UEQ illustrates a sense of satisfaction to its population through assessing different environmental, urban parameters and socio-economic indicators. This chapter elucidates the use of the Geographic Information System (GIS), Principal Component Analysis (PCA) and Geographically-Weighted Regression (GWR) techniques to integrate various parameters and estimate the UEQ of two major cities in Ontario, Canada. Remote sensing, GIS and census data were first obtained to derive various environmental, urban parameters and socio-economic indicators. The aforementioned techniques were used to integrate all of these environmental, urban and socio-economic parameters. Three key indicators, including family income, higher level of education and land value, were used as a reference to validate the outcomes derived from the integration techniques. The results were evaluated by assessing the relationship between the extracted UEQ results

and the reference layers. Initial findings showed that the GWR with the spatial lag model represents an improved precision and accuracy by up to 20% with respect to those derived by using GIS overlay and PCA techniques for the City of Toronto and the City of Ottawa. The findings of the research can help the authorities and decision makers to understand the empirical relationships among environmental factors, urban morphology and real estate and decide for more environmental justice.

4.2 Introduction

The terminology "quality of life" has been continuously discussed in the literature, so as to lay a foundation to serve the subsequent quantification of Urban Environmental Quality (UEQ). Szalai (1980) emphasized that quality of life represents the degree of satisfaction with life and the feeling of well-being, which can be measured by exogenous and endogenous factors. Diener and Suh (1997) concluded the meaning of the quality of life by the satisfaction of life. Raphael *et al.* (1996) further echoed and agreed that quality of life tends more to be the enjoyable degree of a person toward the important responsibilities of his/her life. However, Van Kamp *et al.* (2003) described the quality of life by physical and immaterial equipment, such as health, education, justice, work, family, etc.

UEQ is the consequence of the combination of environmental parameters, including nature, open space, infrastructure, built environment, physical environment amenities and natural resources, and each parameter has its own characteristics and partial quality. Van Kamp *et al.* (2003) addressed that UEQ is an essential part of the quality of life, which has basic concepts, such as health, safety and education, in addition to the physical and environmental parameters. Designing a theoretical framework of UEQ linking to the quality of life is an essential step to understand urban sustainability and human well-being. Such a framework may help to choose the parameters and the integration techniques to evaluate the multidimensional aspects of UEQ. These integration techniques are able to assess the current and predict/estimate the future UEQ, which are modelled by the municipal and city planners (Van Kamp *et al.*, 2003). Thus, the assessment of UEQ can be an efficient tool to provide effective information of urban conditions, sustainable development and regional planning (Faisal and Shaker, 2017). UEQ can be modelled using satellite remote sensing techniques through analysing multi-temporal and multi-resolution data, which are able to give a clear vision for visualizing and understanding the land cover, Land Surface Temperature (LST), water conditions and vegetation in urban areas (Fung and Siu, 2000, 2001). Consequently, several studies in the literatures demonstrated the use of multi-source data to model and assess the UEQ (Nichol and Lee, 2005; Nichol and Wong, 2006; Nichol *et al.*, 2006).

Moore *et al.* (2006) conducted a research study in three U.K. city centre areas, including Clerkenwell in London, Devonshire Quarter in Sheffield and the city centre of Manchester. The main goal of the study was to investigate and understand the UEQ in both subjective and objective bases, which mainly represent the city in mind and the city physically in reality, respectively. The case study divided the project into three sections: (1) outdoor environmental quality, which represents the physical, environmental conditions in the city; (2) perceived environmental quality, which represents the experiences of city residents; and (3) indoor environmental quality, which represents the physical and environmental conditions of residential buildings. Noise levels, carbon monoxide and air temperature were observed over a summer and winter period for the outdoor environmental quality assessment. For the perceived environmental quality, residents from each city were hired to conduct a photo survey and a semi-structured interview to assess residents' experiences within each case study. The levels of carbon dioxide (CO_2), carbon monoxide (CO), thermal conditions (°C) and light intensity were measured for the indoor environmental monitoring. The findings of this case study illustrated the local environmental quality maps and spatial urban environmental factors that represent the environmental quality within the city. The combination of subjective and objective approaches enabled encouraging people to think about how they understand the environment. The proposed method can provide an efficient way for residents worldwide to highlight their concerns, wishes and positive aspects of their local area to support decision makers.

Fobil *et al.* (2011) presented a case study of UEQ in the City of Accra, Ghana. The primary goal of the study was to investigate the relationship between the urban environmental quality and death locations, which was commonly caused by malaria and infectious diarrhoea in low-income countries. First, a total of 65 environmental parameters, such as population and waste generation, water supply and sanitation, hygiene conditions and building structure material, were obtained from the Ghana Census 2000 database. The births and deaths registry in Accra provided the mortality data over the period 1998–2002.

Second, Principal Component Analysis (PCA) was used to integrate the environmental parameters' data and the mortality data of the study area. PCA was used to first compute the correlation among all pairwise parameters. Data reduction was subsequently conducted to reduce the environmental parameters. The results showed that all of the zones were labelled with good, bad or terrible environmental conditions. Third, analysis of variance was used to compare the differences in malaria and diarrhoea mortality levels in the three environmental zones. Fourth, a linear association was conducted between the environmental parameters and malaria and diarrhoea mortalities by using Generalized Linear Models (GLMs). The result demonstrated a strong relationship between environmental parameters and the mortality of malaria. However, there was no strong correlation found between environmental parameters and mortality from diarrhoea. The study illustrated that urban environmental management can be used to reduce the risk of infectious disease in low-income countries.

Lo (1996) introduced the Landsat Thematic Mapper (TM) and social data to assess the quality of life in the city of Athens. The Landsat TM image was first obtained to generate the land use/land cover map and to extract biophysical information from it, including the Normalize Difference Vegetation Index (NDVI) and LST. Socio-economic data were obtained from U.S. census, including population density, per capita income, median home value and percent of college graduates. The maximum likelihood classification was implemented to extract low and high density residential areas, commercial and industrial areas, water, roads, forests and agriculture areas. Principal Component Analysis (PCA) and GIS overlay were used to integrate the land use/land cover, biophysical and socio-economic data. The results showed that NDVI has a strong correlation with per capita income, median home value and percent of college graduates. However, it indicates that NDVI has relatively low correlation with population density, land surface temperature, high density of residential areas, commercial areas and industrial areas. The study showed that the integration of land use/land cover, biophysical and social data can aid in predicting a realistic UEQ for the city.

Another representative study was found in U.S. counties, conducted by (Shoff *et al.*, 2014). The main goal of this case study was to investigate the place-specific risk factors for prenatal care utilization in the U.S. using Spatially-Lagged Geographically-Weighted Regression (GWR-SL). The dependent variable, including late or no prenatal care, was first extracted from the Women's Health Quick Health Data Online from 1999–2001. The late or no prenatal care mainly represents the percentage of women who received prenatal care during their second or third period of pregnancy or did not receive prenatal care at all. The racial composition variables, including the percentage of black females of childbearing age, the percentage of American Indian/Alaskan Native (AIAN) females of childbearing age, the percentage of Asian females of childbearing age and the percentage of Hispanic females of childbearing age were obtained from the above mentioned health data to be included in the analysis. Additionally, the nativity status composition, including the percentage foreign-born, was obtained for the same period and included in the analysis. GWR-SL was implemented in this case study to model the spatial location of prenatal care utilization in U.S. counties. The results of the GWR-SL approach were compared with some of the existing methods, including ordinary least squares and the spatial lag regression model, and the GWR-SL approach showed a better understanding of prenatal care utilization in U.S. counties than the previously mentioned existing approaches. That is mainly because the GWR-SL approach takes into consideration the spatial nature of the data. The findings of this case study help to better estimate and understand the spatial prenatal care utilization in the U.S.

Despite the above successful attempts, the majority of the scholars mainly utilized PCA, GIS analysis or MCE techniques to integrate UEQ parameters (Nichol and Wong, 2009; Fobil *et al.*, 2011; Lo, 1996; Rinner, 2007). The PCA analytical technique has several potential disadvantages: (1) it produces unweighted components, which may not highlight those important parameters; (2) PCA does not work properly in nonlinear relationships; and finally, (3) the minimum number of components is indeterminable (Faisal and Shaker, 2017). The GIS overlay method does not consider correlation among parameters, nor give weight to the parameters. MCE is a weighting process that allows decision makers to modify attribute values of the parameters, which may lead to biased opinions. Numerous researchers (Nichol and Wong, 2009; Liang and Weng, 2011; Rahman *et al.*, 2011; Rinner, 2007) attempted to validate the UEQ results using e-mail questionnaire, field-based questionnaire, interviews with experts and factor analysis. However, these methods can be inaccurate to test the outcomes of UEQ; as a result, it may cause tendentious results. In this research, we attempt to fill several gaps in UEQ research by: (1) utilizing a new method to normalize the UEQ parameters; (2) introducing a new method to weight urban and environmental parameters obtained from diversity data; and (3) proposing a new method to validate urban and environmental parameters with socio-economic indicators for UEQ assessment in two cities in Ontario, Canada.

4.3 Datasets

In this research, the City of Toronto and the City of Ottawa were intentionally selected as case studies due to the data availability and the rapid population growth in these two cities. The datasets used in this study include three broad categories: (1) Landsat TM satellite images; (2) GIS data layers; and (3) socio-economic data. All of the data were collected between the years 2010 and 2011 since GIS data and socio-economic are not consistently available to support the two case studies. A Landsat TM image was downloaded from the United States Geological Survey (2014). The spatial resolution of the Landsat images is 30 m for the multi-spectral bands and 120 m for the thermal band. However, the thermal bands were resampled to a 30-m pixel size from the source of data predominantly to align the thermal band with the multi-spectral bands (Kjaersgaard and Allen, 2009). The image was acquired during the summer season (July) to avoid the appearance of clouds and snow cover.

On the other hand, a total of 14 GIS data layers were acquired from the Scholars GeoPortal (2014) for both cities during the same duration of time. The GIS layer data include land use, population density, building density, vegetation and parks, public transportation, historical areas, Central Business District (CBD), sports area, religious and cultural zone, shopping centres, education institution, entertainment zones, crime rate and health condition. These layers were imported into the ArcGIS platform (ArcGIS, Esri, Redlands, CA, USA) for further analysis. Similar to the remote sensing data, all of the data were projected to the UTM 17 N coordinate system for the City of Toronto and the UTM 18 N coordinate system for the City of Ottawa. Lastly, the socio-economic indicators were derived based on the used census data that were obtained from the census bureau. The census bureau archives hundreds of indicators/information related to socio-economic conditions. In this research, socio-economic indicators, including education (university certificate, diploma or degree), family income and land values, were also obtained for the result validation. Table 4.1 summarizes the data sources being used in this study.

Table 4.1: The data sources for City of Toronto and City Of Ottawa.

City	**Landsat TM**	**GIS Data**	**Census Data**
	Toronto	○ Land Use	
	Path/Row = 18/30	○ Population Density	Socio-economic data were
	Sensor = Landsat TM	○ Building Density	provided by the census bureau.
	Date = 23 June 2011	○ Vegetation and Parks	Socio-economic data:
		○ Public Transportation	
Toronto	Ottawa	○ Historical Areas	○ Education
	Path/Row = 16/28	○ Central Business Districts (CBD)	○ Family Income
Ottawa	Sensor = Landsat TM	○ Sports Areas	○ Land Values
	Date = 11 September 2011	○ Religious and Cultural Zones	
		○ Shopping Centres	
	Remote sensing data:	○ Education Institutions	
		○ Entertainment Zones	
	○ LST	○ Crime Rate	
	○ NDVI	○ Health Condition	
	○ NDWI	○ Areas Close to Water Bodies	
	○ NDBI and Built-up Area		

4.4 Methodology

Figure 4.1 shows the overall workflow for the two case studies (the City of Toronto and the City of Ottawa), which can be summarized by the following steps. The Landsat images were imported into PCI Geomatics V10.1(*Geomatica*, version 10.1, PCI Geomatics, Markham, ON, Canada, 2007), clipped and then projected into the UTM coordinate system. The absolute atmospheric correction model, ATCOR2 (Atmospheric Correction and Haze Reduction), built-in PCI Geomatics software was used to compute the results for several bio-physical parameters (NDVI, NDWI, built-up index and LST) (Richter, 1998). ATCOR2 was utilized to first perform radiometric calibration and to remove the effects that change the spectral characteristics of the land features (Paolini *et al.*, 2006). Sensor parameters, including sensor type, acquisition date, Sun elevation, Sun zenith and pixel size, were obtained in addition to weather conditions (air temperature and visibility) to conduct the subsequent atmospheric correction. The calibration parameters for Landsat TM sensor (biases and gains) were also incorporated into the atmospheric correction, as is described in (Richter, 1990). In this research, biophysical parameters, including NDVI, NDWI, built-up index and LST, were derived from the Landsat images. Urban, environmental parameters and socio-economic indicators were all derived from GIS and census data to combine all of the parameters together for further analysis. The methodological contribution of this research work is to implement the GIS overlay, PCA and GWR (ordinary GWR, GWR with spatial error model and the GWR with spatial lag model) to integrate all urban, environmental parameters and socio-economic indicators. Then, socio-economic indicators, including family income, higher education level and land values, were investigated to validate the final outcomes from the integration methods. The evaluation of the binary classifiers algorithm was performed to assess the precision and accuracy of each integration method. Based on the precision and accuracy of the integration methods, the optimal integrated method can be determined to estimate the best UEQ location in the two case studies.

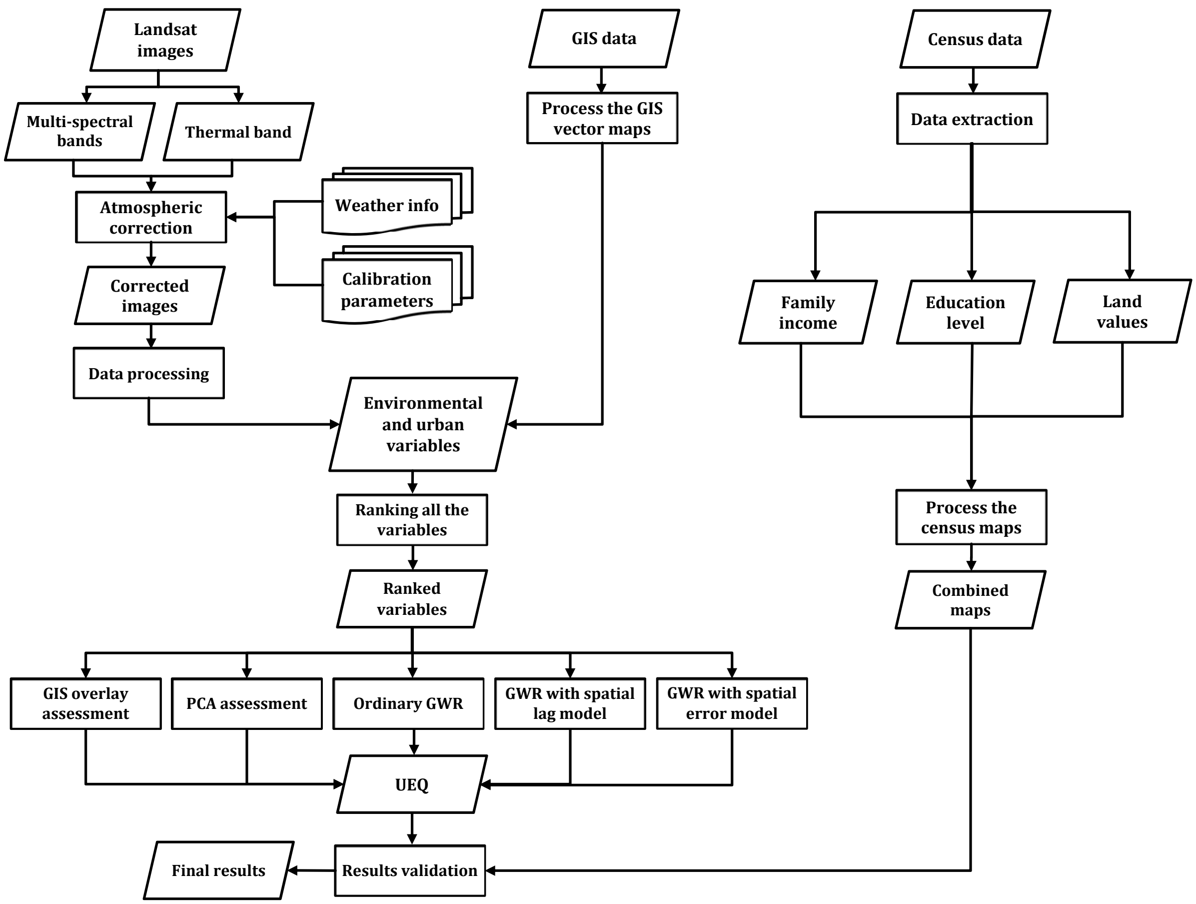


Figure 4.1: The overall workflow for improving the accuracy of UEQ assessment.

4.4.1 Ranking the Parameters

Since the parameters as mentioned earlier were extracted from different data sources, they may have different scale levels and cannot be combined into a particular unit. Therefore, all of the obtained data (parameters), including raster, census and GIS data, were first transformed into one scale (sub-neighbour), as shown in Figure 4.2. To standardize the parameters and represent the significant level of each polygon in the parameter, the Z-score method was performed for all of the parameters. The Z-score model is a statistical measurement that is able to standardize a wide range of data to represent the significant changes across data Cheadle *et al.* (2003).

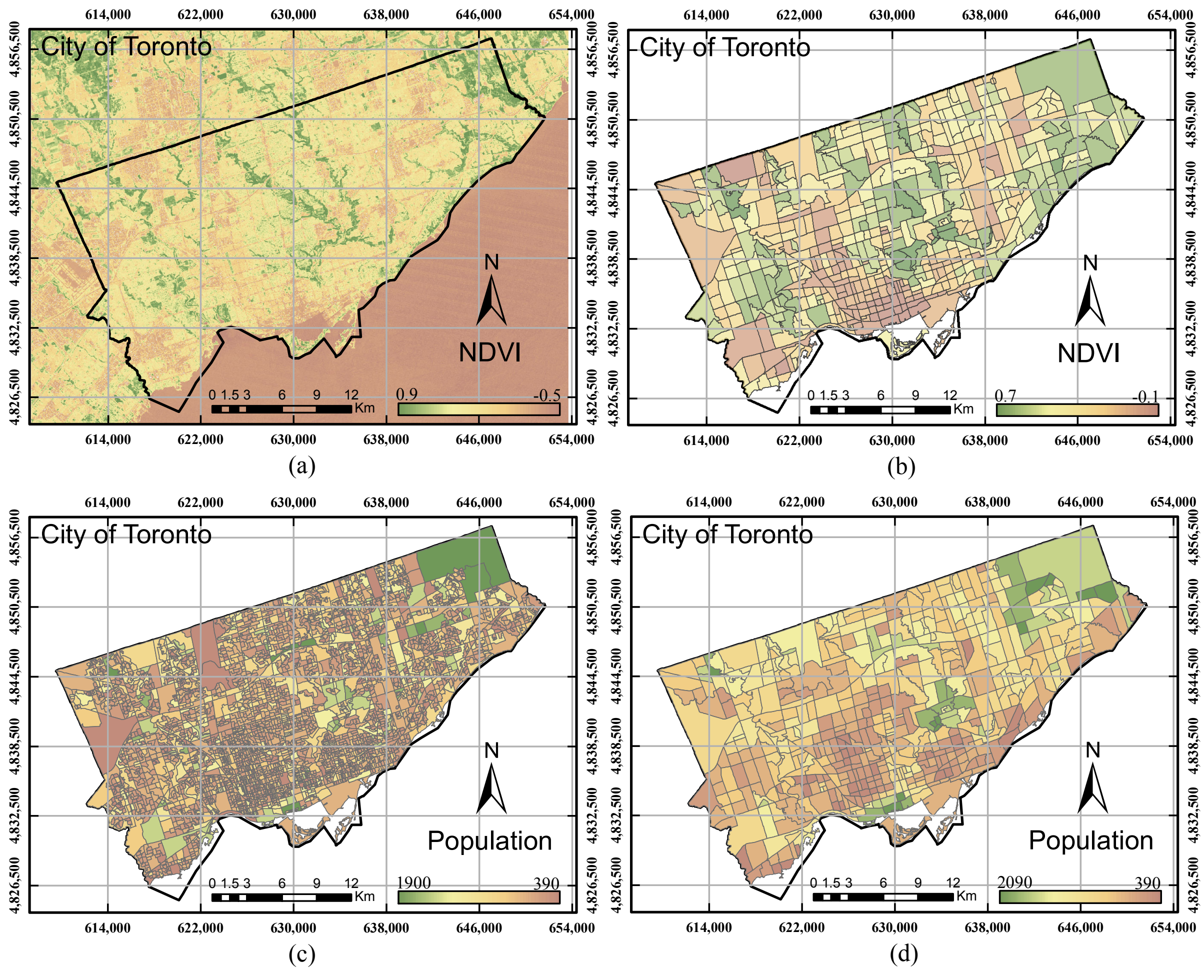


Figure 4.2: (**a**) NDVI image derived from the Landsat image (raster data); (**b**) NDVI map after transformation (vector data); (**c**) population layer at the census tract level; (**d**) population layer after transformation to the sub-neighbour level.

The following Equation (4.1) shows the first step to normalize the parameters using the Z-score:

$$Z_i = \left[\frac{x_i - \mu}{\sigma}\right] \tag{4.1}$$

where x is the observation values (polygons), i is the parameter, μ is the mean value of the parameter and σ is the standard derivation of the parameter. The second step is to use linear interpolation to rank the parameters from 1–10. The polygon within the parameter that has a high Z-score number will represent high values, for example 10. The polygon that has a low Z-score will result in a value of 1. However, for those parameters having negative relationships with respect to UEQ, such as crime rate, industrial areas, LST, etc., these parameters are inversely presented (e.g., the highest LST will take a value of 1, and the lowest LST value will get 10), as shown in Figure 4.3. The following Equation (4.2) shows how linear interpolation was calculated:

$$Rank = \left[\frac{(Obs - Obs_{max})(Rank_{min} - Rank_{max})}{(Obs_{min} - Obs_{max})}\right] + Rank_{max} \tag{4.2}$$

where Obs is the current observation value, Obs_{max} is the maximum observation value, Obs_{min} is the minimum observation value, $Rank_{max}$ is the maximum ranking value, $Rank$ is the determined ranking value and $Rank_{min}$ is the minimum ranking value.

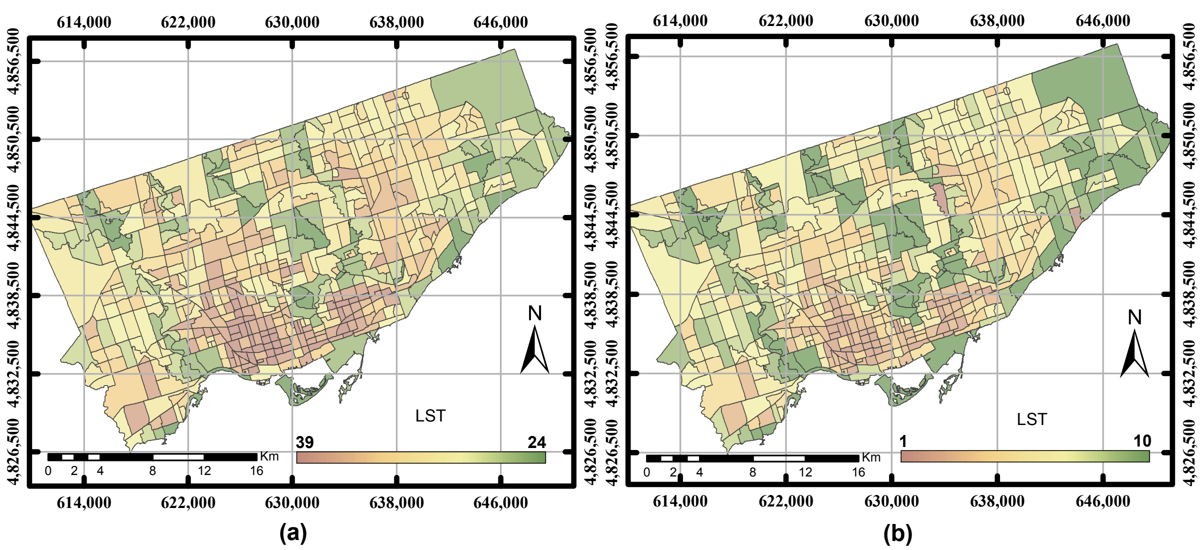


Figure 4.3: (**a**) The LST layer in degrees Celsius before ranking the parameter; (**b**) the ranking LST after the normalization.

4.4.2 Data Integration of Multiple Environmental and Urban Parameters

Integration techniques can be used to combine remote sensing and GIS data for urban modelling and analysis (Weng, 2002). Previous studies demonstrated two integration methods, mainly PCA and GIS overlay, which are able to combine various parameters from a diverse source of data. In this research work, three approaches were demonstrated to integrate the aforementioned environmental and urban parameters. These two existing approaches (PCA and GIS overlay) were first implemented, and subsequently, we investigated the use of GWR techniques (ordinary GWR, the GWR with spatial lag model and the GWR with spatial error model) to integrate all of the aforementioned parameters, which can lead to an improved estimation of UEQ.

Geographic Information System Overlay

GIS overlay is a multi-criteria application that uses data layers for specific environmental thresholds. Remote sensing data are commonly presented as digital data in raster format. However, census data are usually stored in GIS vector format. Remote sensing data can thus be integrated with socio-economic data by converting remote sensing data from raster to vector data (Li and Weng, 2007). In this research work, the GIS overlay integration method was used to combine the urban and environmental parameters to serve for the UEQ assessment. After, we transform all of the obtained data into sub-neighbours and rank the parameters from 1–10 using Equations (4.1) and (4.2). The sum of the data layers can thus illustrate the result of UEQ.

Principal Component Analysis

PCA is an analysis technique that compresses high dimensional data into a small size of data and retains most of the variance of the data (Jensen, 2005). PCA is commonly used in many remote sensing applications. The covariance matrix of standard PCA may not be the best option for data that have different measurement units. The correlation matrix can be used instead of the covariance matrix to standardize each parameter to the variance unit or zero means. In this research work, two case studies were conducted to assess the UEQ in the City of Toronto and the City of Ottawa, respectively. The observation values of the GIS polygons of each parameter were employed in the PCA model to determine the UEQ, as shown in Figure 4.4.

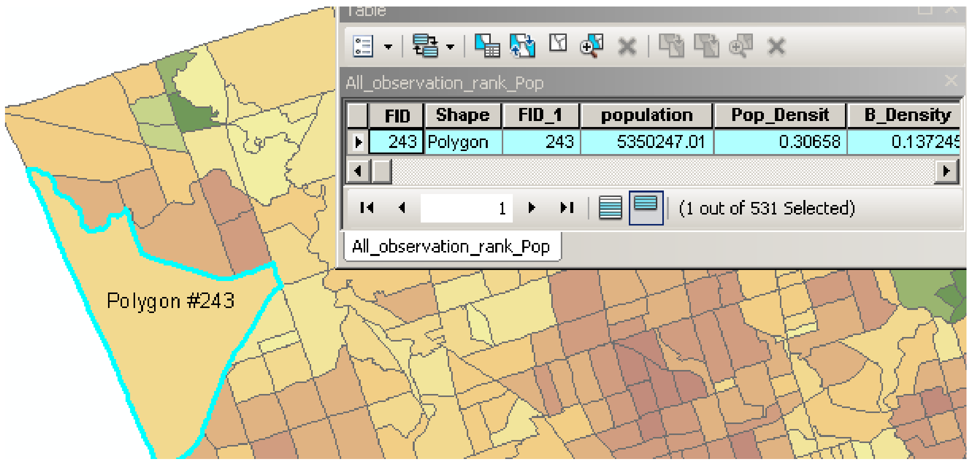


Figure 4.4: The GIS polygons of the parameters.

PCA can be computed by determining the eigenvectors and eigenvalues of the correlation matrix. The first step to compute PCA is to calculate the correlation matrix. The correlation of two random parameters can be computed by using the following Equation (4.3):

$$r_{y1,y2} = \left[\frac{cov_{(y_1 i, y_2 i)}}{\sigma_{(y_1 i)} . \sigma_{(y_2 i)}} \right] \tag{4.3}$$

where r is the correlation matrix for parameters y_1 and y_2, respectively, $cov_{(y_1 i)}$ and $cov_{(y_2 i)}$ are the covariance matrix for parameter y_1 and y_2, respectively, and $\sigma_{(y_1 i)}$ and $\sigma_{(y_2 i)}$ are the standard deviation for parameter y_1 and y_2, respectively, at location i.

The second step is to calculate the eigenvalues of the correlation matrix. The eigenvalue measures the scale of the data. The parameters that have eigenvalues greater than one will be a good rule of thumb to represent most of the variance of the data (Cliff, 1988). Eigenvalues can be computed by using the following Equation (4.4):

$$\det(A - \lambda I) = 0 \tag{4.4}$$

where A is the correlation, λ is the eigenvalues and I is an N by N identity matrix.

The third step is to calculate the eigenvector of the correlation matrix. The eigenvector measures the direction of the data. Eigenvectors can be computed by using the following Equation (4.5):

$$(A - \lambda I)X = 0 \tag{4.5}$$

where A is the correlation matrix, λ is the eigenvalues and X is the eigenvector.

The new *Obs* (observation number) for the new image can be determined using the following Equation (4.6) (Jensen, 2005):

$$New(Obs_i) = \sum_{i=1}^{n} a_{kp} * Obs_i \tag{4.6}$$

where a_{kp} is the eigenvector for parameter k component p and *Obs* is the observation number in polygon i.

Ordinary Geographically-Weighted Regression

One of the limitations of using PCA is that it produces unweighted components. GWR can be used to weight the spatial location of each parameter. The dependent parameter indicates the UEQ outcome, which was derived from GIS overlay method. That is mainly because the GIS overlay was found to be more emblematic for UEQ in some previous studies and one of our parallel studies (Nichol and Wong, 2009; Faisal and Shaker, 2017). The independent parameters are the urban and environmental parameters, which were derived from the remote sensing and GIS data, such as population density, building density, NDVI, public transportation, etc. The weight can be given to some location based on the nearness and similarity of the estimated parameters at some location. Thus, the observations that are located nearer to the estimated location would have a higher weight. However, the observations that are located far from the estimated location would have a lower weight. Assume we have a dataset that consists of a dependent variable y and a set of independent variables $(x_1, x_2, x_3 ... x_n)$, and for each of the i observations in the dataset, a measurement of its position is available in a suitable coordinate system (Charlton *et al.*, 2009). Equation (4.7) shows the ordinary GWR model:

$$y_i = a_{1i}x_1 + a_{2i}x_2 + a_{3i}x_3 \; .... + a_{ni}x_n \tag{4.7}$$

where $a_{1i}...a_{ni}$ refer to the coefficients that define a spatial relationship with respect to its surroundings at location i. The outcomes of y_i indicates a new dependent variable if we have the dataset of the independent variables x at location i. The GWR mathematical model thus considers the weights with respect to the surroundings at location i to estimate coefficients $a_{1i}...a_{ni}$ that define a spatial relationship with respect to its surroundings at location i. The following form (4.8) represents the coefficients (a_i) at location i:

$$a_i = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{Y} \tag{4.8}$$

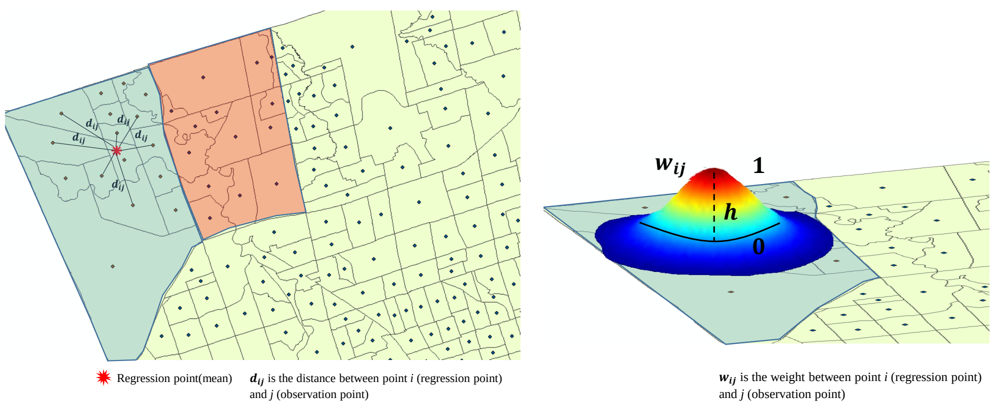
where $\mathbf{W}_i$ is a square matrix of weights relative to the position of i in the study area; $\mathbf{X}$ is the independent variables matrix; and $\mathbf{Y}$ is the dependent variable. The $\mathbf{W}_i$ matrix captures dependency relations between the observations, which represent the geographical weights in the diagonal and 0 in its off diagonal matrix Borst and McCluskey (2007).

In this research work, the distance-based weights algorithm was implemented to create the diagonal weighted matrix. This method can be used to avoid non-weighted isolated polygons and polygons that are located inside other polygons. An optimum bandwidth can be defined through using certain techniques, including the Cross-Validation (CV) and Akaike Information Criterion (AIC), to derive the goodness of fit (Lu *et al.*, 2014). However, numerous researchers suggested different kernel functions to derive the bandwidth, such as the distance based on the taxicab geometry (Krause, 1987), the chamfer distance designed for a lattice or grid space (Leymarie and Levine, 1992; De Smith, 2004), the shortest path distance (Smith, 1989) and the qualitative distance by translating an absolute distance metric to linguistic terms (Guesgen and Albrecht, 2000; Yao and Thill, 2005). In this study, the first step to compute the weighted matrix is to determine the neighbours, mainly based on the k-nearest neighbour weighted method. For instance, we could generate centre points of a 10 $\times$ 10 lattice as a mean point location or regression point to measure the distances, as shown in Figure 4.5a. In addition, the polygons can be computed based on a weighting scheme known as a kernel, and in this study, we used the Gaussian shape kernel, as shown in Figure 4.5b. The following form (4.9) represents the weighting scheme for the distance-based method (Lu *et al.*, 2014;

Fotheringham *et al.*, 2003).

$$w_{ij} = \exp\left[-0.5(d_{ij}/h)^2\right] \tag{4.9}$$

where w_{ij} is the spatial weight between observation point j and regression point i, d_{ij} is the distance between observation point j and regression point i and h is the kernel bandwidth defined by the distance between the regression location and the k-th nearest observation.



(**a**) (**b**)

Figure 4.5: Weighted distance method. (**a**) k-nearest neighbour; (**b**) Gaussian shape kernel.

Geographically-Weighted Regression with Spatial Lag Model

The spatial lag model is one of the dominant spatial autoregressive regression models that has been used in many research studies (Ward and Gleditsch, 2008; Páez *et al.*, 2002). Shoff *et al.* (2012) used the GWR with spatial lag approach to model and predict the U.S. prenatal care utilization at the county level dataset. The spatial lag model essentially heals spatial heterogeneity by including an autocorrelation coefficient and spatial weight matrix in the weighted regression model. The SLM is expressed as the following Equation (4.10):

$$\mathbf{Y}_i = \rho_i \mathbf{W}_i y_i + \mathbf{X}_k \beta_i + \epsilon \tag{4.10}$$

where $\mathbf{Y}$ is an N by 1 vector of observations on the dependent variable, i is the location coordinates (centroid of the county), $\mathbf{W}$ is an N by N specifying spatial weights matrix, which indicates the distance relationship between locations i and j, and ρ_i is the spatial lag dependence between county level percentages of UEQ at location i. For a given location, say j, ρ indicates the relationship between j's dependent variable (UEQ) and the dependent variable of j's neighbours defined by the distance weight matrix. Positive ρ refers to a positive spatial autocorrelation; and if ρ is negative, then negative spatial autocorrelation is determined. β_i is a K by 1 vector of regression coefficients associated with $\mathbf{X}_k$ at location i. ϵ is an N by 1 vector of the error term.

Geographically-Weighted Regression with Spatial Error Model

The GWR with spatial error model is appropriate when we are interested in correcting spatial autocorrelation due to the use of spatial data. In this case, the structure or spatial heterogeneity of the spatial relationship is missing. Therefore, we include the spatial autoregressive error term due to unobservable features or omitted variables that are related to locations (Anselin and Bera, 1998). The GWR with spatial error model is expressed as the following Equations (4.11) and (4.12):

$$\mathbf{Y}_i = \mathbf{X}_k\beta_i + \epsilon \tag{4.11}$$

$$\epsilon = \lambda_i\mathbf{W}_i\epsilon + \zeta \tag{4.12}$$

where λ_i is the spatial autoregressive coefficient for the error lag $\mathbf{W}_i\epsilon$ and ζ is a vector of independent identically distributed errors.

Accuracy Assessment

Data validation is one of the major concerns in UEQ research work. Several researchers attempted to assess the accuracy of the UEQ results using different methods, including e-mail questionnaire, field-based questionnaire, asking experts and factor analysis. Regardless of the considerable amount of e-mail surveys or field-based questionnaires, both approaches are time consuming and budget dependent. Besides, factor analysis used in the previous work was performed using the same parameters that have been incorporated to compute the UEQ, which make it unreliable and biased. Numerous UEQ studies did not perform any field survey or even results validation (Lo, 1996; Moore *et al.*, 2006; Fobil *et al.*, 2011). On the other hand, some of the literatures highlighted a high correlation between socio-economic indicators including (university certificate or diploma, family income and land values) and the quality of living (Becker and Mulligan, 1997; Kahn, 2007; Reginster and Goffette-Nagot, 2005; Din *et al.*, 2001). Since there is a lack of ground reference to validate the results in this study, we propose to use these socio-economic indicators for data validation and to assess the UEQ results. All observed data of the three socio-economic indicators were normalized to be in the same scale from 1–10. Then, the sum of the socio-economic indicators can thus present the result of the reference, as shown in Table 4.2.

Table 4.2: The sum of the socio-economic indicators.

Polygon ID	**Income**	**Education**	**Land Value**	**Reference Layer**
1	8	5	7	20

The first step to validate the results is to extract the observation's values that are higher than the mean in each parameter and reference layer. That is mainly because in this study, we need to highlight the higher UEQ areas. Second, the evaluation of binary classifiers approach was used to evaluate the UEQ based on the following two performance measures through data interpretation: *Precision* and *Accuracy* (Powers, 2011).

$Precision(P)$ is a measure that evaluates the probability that a positive outcome is correct using Equation (4.13):

$$P = \frac{\mid TP \mid}{\mid TP \mid + \mid FP \mid} \tag{4.13}$$

$Accuracy(Acc)$ evaluates the effectiveness of the classifier by its percentage of correct predictions using Equation (4.14):

$$Acc = \frac{\mid TN \mid + \mid TP \mid}{\mid FN \mid + \mid FP \mid + \mid TN \mid + \mid TP \mid} \tag{4.14}$$

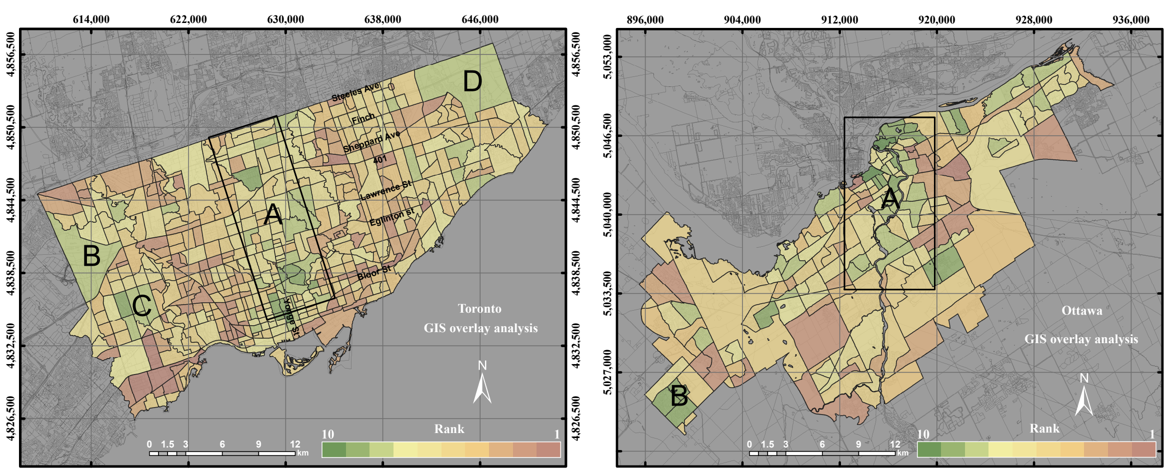
where TP refers to "True Positive", which means the above mean polygons derived from the proposed method are being matched physically in the reference layer; TN refers to

"True Negative", which represents the above mean polygons that are not detected in the proposed method and the reference layer; FP refers to "False Positive", which means the above mean polygon derived from the proposed method does not really exist in the reference layer; and FN refers to "False Negative", which means the above mean reference polygons do not exist in the proposed method. With these three indicators, we assessed the UEQ layer from the results of each proposed method, including GIS overlay, and PCA assessed the best method for our datasets.

4.5 Results and Discussion

4.5.1 GIS Overlay Analysis

Figure 4.6a shows the UEQ derived in the City of Toronto using the GIS overlay. The distribution of UEQ in the City of Toronto shows that the highest UEQ zones were found in areas (A, B, C and D) in green colour, while the lowest UEQ zones are indicated as brown colour in the city.



(**a**) (**b**)

Figure 4.6: The Urban Environmental Quality (UEQ) derived using the GIS overlay method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa.

The highest UEQ zones are the consequence of the summation of all of the positive parameters that are located within Zones A–D. However, negative values of the parameters, including crime, industrial areas and high LST, are consistently located in the brown zones within the city. In contrast, the highest values of UEQ areas were found in the high and moderate density areas, while the lowest values were found in the industrial and low density areas. Figure 4.6b shows the UEQ derived in the City of Ottawa using the same method. Apparently, the distribution of UEQ in the City of Ottawa showed that the highest UEQ zones were found in Zones A and B. These areas are mainly located in the down town of the city and the new urban area in Zone B. The highest values of UEQ areas were consistently found in the high and moderate density areas. However, some suburban areas located in Zone B showed high UEQ values, and that could be due to the increase of income of the household, resulting in a move to the suburbs, especially in automobile-dependent cities, such as the City of Ottawa (Kahn, 2007; Turcotte, 2008).

4.5.2 Principal Component Analysis

Table 4.3 represents the correlation coefficient matrix among all of the parameters in the City of Toronto. Population density reported a moderate positive correlation coefficient with historical areas parameter (0.66).

Table 4.3: The correlation coefficient matrix among all of the parameters derived from the PCA method in the City of Toronto.

	PD	**BD**	**PT**	**Veg**	**NDVI**	**NDWI**	**BU**	**LST**	**H**	**Ind**	**CBD**	**Sc**	**Ent**	**He**	**Rel**	**SP**	**Sea**	**CR**	**SH**
PD	1.00	0.34	0.14	−0.14	−0.11	0.11	0.12	0.12	0.66	−0.04	0.08	−0.17	−0.02	0.03	−0.11	−0.04	−0.06	0.02	−0.04
BD		1.00	0.40	−0.61	−0.68	−0.67	0.67	0.78	0.44	0.07	0.39	−0.05	0.14	0.11	0.16	0.02	0.21	0.22	0.05
PT			1.00	−0.37	−0.37	−0.36	0.38	0.46	0.12	0.15	0.16	−0.09	−0.04	−0.01	0.05	−0.03	0.12	0.12	0.04
Veg				1.00	0.66	0.55	−0.56	−0.66	−0.11	−0.13	−0.09	−0.03	0.05	−0.03	−0.13	0.03	−0.30	−0.11	−0.02
NDVI					1.00	0.88	−0.90	−0.80	−0.30	−0.37	−0.37	0.02	−0.27	−0.10	−0.29	−0.09	−0.27	−0.35	−0.23
NDWI						1.00	−0.89	−0.77	−0.31	−0.39	0.37	−0.02	0.29	0.11	0.31	0.10	0.25	−0.35	0.26
BU							1.00	0.79	0.30	0.50	0.35	−0.01	0.27	0.10	0.31	0.09	0.27	0.35	0.24
LST								1.00	0.18	0.19	0.25	−0.02	0.05	0.05	0.14	0.00	0.31	0.19	0.06
H									1.00	−0.01	0.50	−0.05	0.43	0.24	0.09	0.16	−0.05	0.33	0.19
Ind										1.00	0.03	0.02	0.08	−0.01	0.31	0.05	0.06	0.12	0.14
CBD											1.00	−0.05	0.37	0.19	0.07	0.09	−0.07	0.38	0.16
Sc												1.00	0.04	0.12	0.25	0.05	0.21	0.00	0.03
Ent													1.00	0.30	0.26	0.39	0.00	0.38	0.49
He														1.00	0.30	0.49	−0.03	0.21	0.38
Rel															1.00	0.44	0.11	0.15	0.41
SP																1.00	0.02	0.18	0.62
Sea																	1.00	0.01	0.03
CR																		1.00	0.27
SH																			1.00

PD: Population Density; BD: Building Density; PT: Public Transportation; Veg: Vegetation Areas; BU: Built-up Areas; LST: LST; H: Historical Areas; Ind: Industrial Areas; Sc: School Areas; Ent: Entertainment Areas; He: Health Condition; Rel: Religion Areas; SP: Sport Areas; Sea: Areas Close to a Water Body; CR: Crime Rate Areas; SH: Shopping Areas.

Where building density showed a moderate negative correlation with green vegetation

(−0.61), NDVI (−0.68), NDWI (−0.67) and a positive correlation with built-up areas (0.67) and LST (0.78). NDVI exposed a strong positive relationship with NDWI (0.88) and a moderate negative correlation with green vegetation (0.66). However, NDVI demonstrated a high negative correlation with the built-up areas parameter (−0.90) and LST (−0.80) and also revealed a moderate negative correlation with building density (−0.68). The built-up areas parameter reported a strong positive correlation with building density (0.67) and LST (0.79). The built-up areas parameter revealed a negative correlation with NDVI (−0.90) and NDWI (−0.89). NDVI stated a very high correlation with NDWI and a negative correlation with the built-up areas parameter and LST. NDVI also demonstrated a moderate negative correlation with building density, which indicates that high NDVI values represent low LST and low high building density areas with more green areas.

In the City of Ottawa, the building density parameter reported a moderate negative correlation coefficient with green vegetation (−0.61), NDVI (−0.66) and NDWI (−0.64) and a positive correlation with built-up areas (0.64) and LST (0.73).

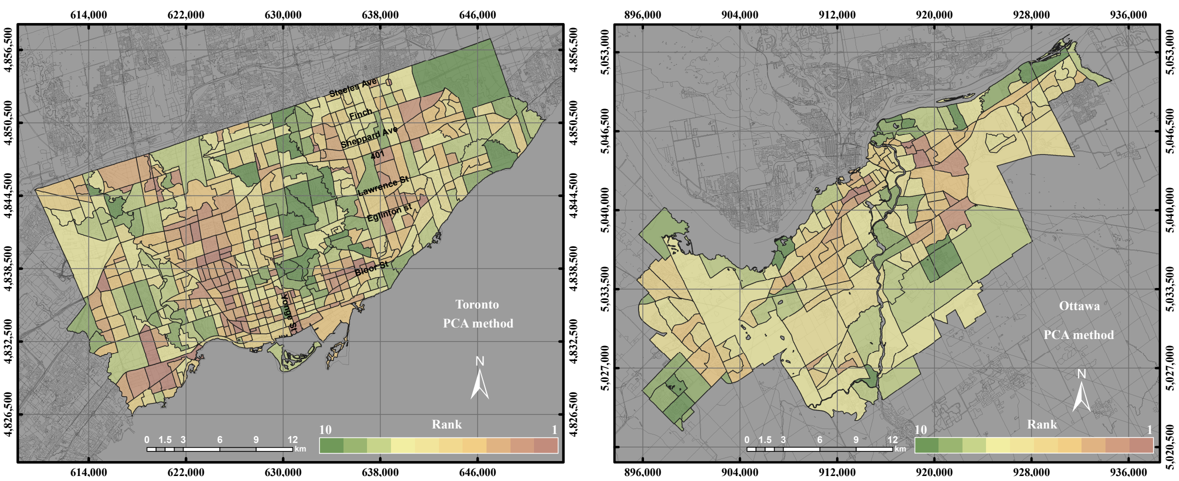
Table 4.4: The correlation coefficient matrix among all of the parameters derived from the PCA method in the City of Ottawa.

	PD	**BD**	**PT**	**Veg**	**NDVI**	**NDWI**	**BU**	**LST**	**H**	**Ind**	**CBD**	**Sc**	**Ent**	**He**	**Rel**	**SP**	**Sea**	**CR**	**SH**
PD	1.00	0.46	0.29	−0.36	−0.28	−0.25	0.28	0.36	0.36	0.30	0.25	0.35	0.06	−0.02	−0.27	−0.18	−0.05	0.13	−0.11
BD		1.00	0.41	−0.61	−0.66	−0.64	0.64	0.73	0.22	0.45	0.38	0.06	0.14	0.13	0.20	−0.02	0.24	0.22	−0.05
PT			1.00	−0.31	−0.25	−0.24	−0.23	−0.37	−0.36	0.33	0.57	−0.05	0.12	0.03	0.02	0.01	0.36	−0.13	0.06
Veg				1.00	0.57	0.56	−0.57	−0.68	−0.17	−0.17	−0.10	0.04	0.04	0.01	−0.14	−0.09	−0.30	0.11	−0.01
NDVI					1.00	0.97	−0.95	−0.77	0.16	−0.40	−0.37	0.03	−0.27	−0.10	−0.30	0.01	−0.29	0.36	−0.24
NDWI						1.00	−0.96	−0.75	−0.24	−0.39	−0.34	−0.03	0.28	0.09	0.32	0.01	0.29	−0.35	0.25
BU							1.00	0.77	−0.16	0.55	0.51	0.02	−0.24	−0.09	−0.30	−0.01	−0.30	0.33	−0.23
LST								1.00	0.21	0.29	−0.23	0.02	−0.04	−0.05	−0.20	0.03	−0.35	0.23	−0.07
H									1.00	−0.29	0.43	0.02	−0.04	−0.04	−0.19	0.02	−0.34	0.21	−0.07
Ind										1.00	0.78	−0.07	0.47	0.23	0.11	0.01	−0.02	−0.33	0.17
CBD											1.00	−0.06	0.38	0.19	0.07	−0.01	−0.06	−0.36	0.15
Sc												1.00	0.04	0.10	0.23	0.14	0.21	0.02	0.04
Ent													1.00	0.27	0.19	0.18	0.01	−0.27	0.41
He														1.00	0.17	0.07	−0.03	−0.15	0.19
Rel															1.00	0.19	0.17	−0.11	0.25
SP																1.00	0.01	0.05	0.23
Sea																	1.00	−0.08	0.04
CR																		1.00	−0.12
SH																			1.00

The green areas parameter also exposed a moderate negative correlation with LST. The data derived from remote sensing data, including NDVI, NDWI, the built-up areas parameter and LST, have high to moderate correlation with each other. NDVI has a high positive correlation with NDWI (0.97) and a high negative correlation with the built-up areas parameter (−0.95). However, NDVI established a moderate negative correlation with LST (−0.77). LST also showed a moderate negative correlation with the green areas parameter (−0.68) and NDWI (−0.75), but a moderate negative correlation with the built-up areas parameter (0.77). The industrial areas parameter revealed a notable moderate positive correlation with CBD (0.78) as shown in Table 4.4. In addition, these observations determined that the above-mentioned remote sensing parameters represented high to moderate correlation among each other. The results also indicated that there are some industrial areas located close to the down town zone that may affect the final results of the UEQ. As mentioned in Section 4.4.2, data reduction can improve the data processing and cost. Therefore, the PCA approach was used to reduce the size of the data.

In this study, five components were extracted in the PCA approach for the City of Toronto, which have an eigenvalue greater than one, as shown in Figure 4.7(a). The total variance of the five components is 75% of the overall variance of the data. The preliminary analysis revealed that Component 1 has 36% of the total variance of the dataset. Component 1 shows strong positive loadings with NDVI (0.88), NDWI (0.86), building density (0.80), LST and historical areas (0.86) and strong negative loadings with LST (−0.86) and built-up areas (−0.86). In addition, Component 1 is the best to represent the green areas within the city. Component 2 reveals about 16% of the dataset, which mainly represents industrial areas with a positive correlation of 0.63 and CBD with a positive correlation of 0.76. Component 2 can be used to describe more about the urban areas. Component 3 represents 9% of the dataset, which mainly represents only sports areas with a positive correlation of (0.81). Component 4 reveals 7% of the dataset, which basically represents public transportation with a positive correlation of 0.70. The final map has a higher correlation (0.7) with the combination of Components 1 and 2. Such findings can reveal that the parameters, which are represented in Components 1 and 2, can be used to reveal the UEQ within the city.

In the City of Ottawa, six components were extracted in the PCA approach that has an eigenvalue larger than one, as shown in Figure 4.7(b). The outcome revealed that Components 1 and 2 have 56% and Components 3, 4, 5 and 6 have 21.1% of the total variance of the dataset. The results showed that Component 1 was highly correlated with NDVI (0.88), NDWI (0.86) and green vegetation (0.80) and has a strong negative correlation with LST (−0.84) and built-up areas (−0.86). Similar to the City of Toronto case study, Component 1 can be used predominantly to derive the green areas within the City of Ottawa. On the other hand, Component 2 detects about 18.4% of the dataset, which mainly represents industrial areas with a positive correlation of 0.72, CBD with a positive correlation of 0.66 and LST with a positive correlation of 0.70. Component 2 can be used to represent more about the urban areas. The findings of the City of Ottawa case study can reveal that the parameters, which are described in Components 1 and 2, can be used to represent the UEQ within the city.



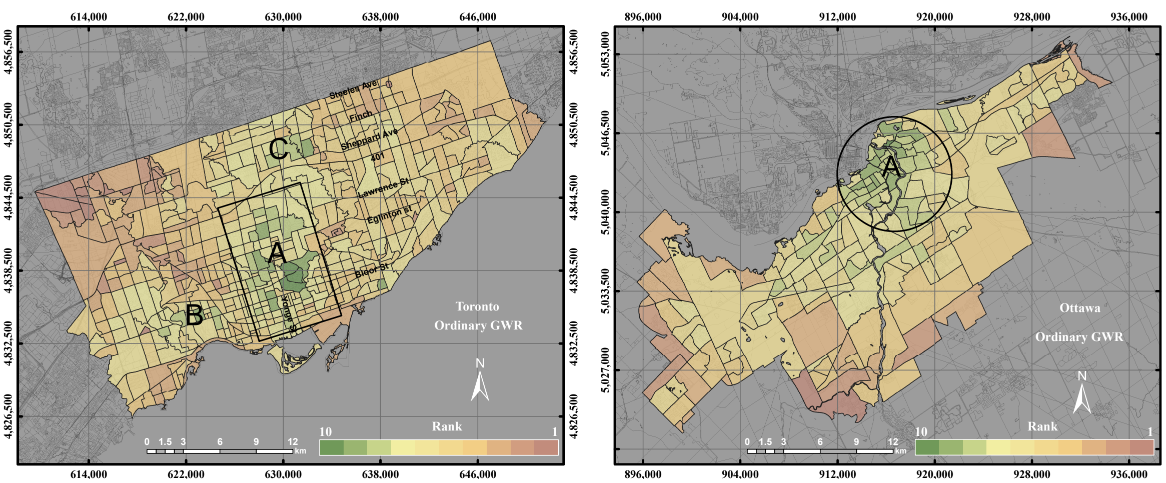
(**a**) (**b**)

Figure 4.7: The UEQ derived using the PCA method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa.

4.5.3 Geographically-Weighted Regression

Ordinary Geographically-Weighted Regression

Figure 4.8(a) shows the derived UEQ using the ordinary GWR in the City of Toronto. The distribution of UEQ in the City of Toronto showed that the highest UEQ zones were mainly found in Zones A, B and C displayed in green colour, which are mainly located in the north and middle of town, as well as the west of the city. However, the lowest UEQ zones are located in the northwest and northeast of the city. The ordinary GWR investigates the spatial weight with respect to the city's polygons and its surrounding polygons. Thus, the outcome showed the highest UEQ zones clustered in the middle, north and west of the city. The highest UEQ values can be ascribed by all of the positive parameters as previously mentions in the result of the GIS overlay. Figure 4.8(b) reveals the derived UEQ for the City of Ottawa using the ordinary GWR. The distribution of the higher values mainly is located in the city centre and the middle of town, as shown in Zone A in Figure 4.8(b). The lowest UEQ zones mostly are located in the remote areas of the city. That could be again because the City of Ottawa is not a high dense city, and many positive parameters are located in the down town and middle of town of the city.

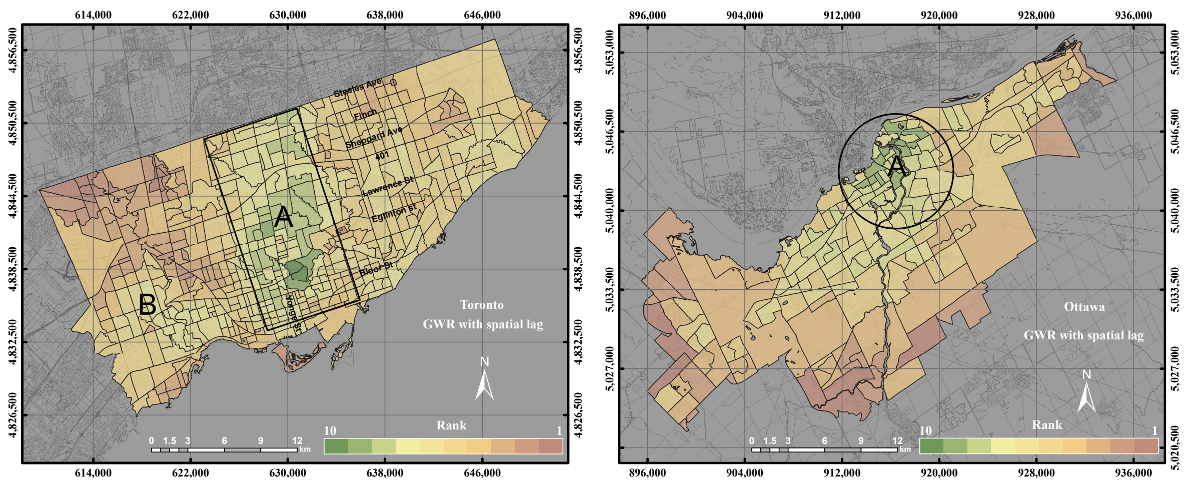


(**a**) (**b**)

Figure 4.8: The UEQ derived using the ordinary GWR method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa.

Geographically-Weighted Regression with Spatial Lag Model

Figure 4.9(a) shows the derived UEQ using the GWR with spatial lag model in the City of Toronto. The distribution of UEQ in the City of Toronto shows that the highest UEQ zones were found in Zone A and Zone B with respect to the UEQ values within the city, while those UEQ zones with low values were located in the northwest and northeast of the city. Since the spatial lag model mainly heals the spatial heterogeneity by including an autocorrelation coefficient and spatial weights matrix in the weighted regression, thus the outcome of the spatial lag model clustered the highest UEQ zones in the middle and north of town of the city as shown in Zones A and B. That is mainly because all of the positive parameters, including (high vegetation areas, historical areas, areas supported by public transportation, low crime rate, etc.), are officially located within Zones A and B. In the City of Ottawa, the results of the GWR with spatial lag model revealed high UEQ values in the city down town and middle of town, as shown in Figure 4.9(b). On the other hand, the lowest UEQ values are located in the suburban areas where there is a lack of public transportation, schools, hospitals and city activity.



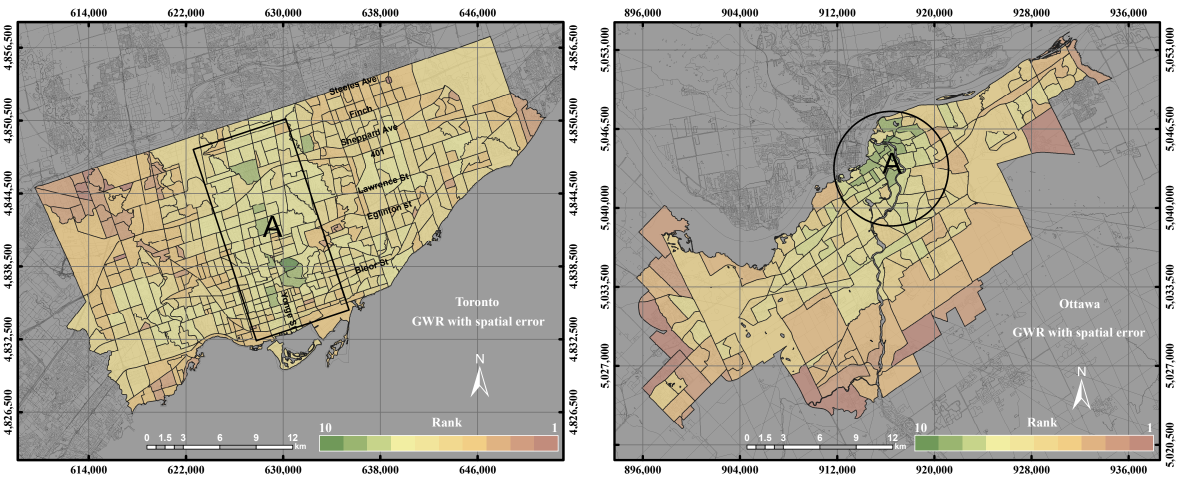
(**a**) (**b**)

Figure 4.9: The UEQ derived using the GWR with spatial lag method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa.

Geographically-Weighted Regression with Spatial Error Model

The distribution of UEQ derived from using the GWR with spatial error model in the City of Toronto shows that the highest UEQ zones were clustered on Yonge Street, as shown in Figure 4.10(a). The lowest UEQ zones are also indicated in the northwest and northeast of the city. GWR with spatial error is able to correct the spatial autocorrelation of spatial data. Thus, the outcome shows the highest UEQ zones located on the most active street within the City of Toronto. That is mainly because most of the positive parameters are located along Yonge Street. Figure 4.10(b) revealed the distribution of UEQ derived from using the GWR with the spatial error model in the City of Ottawa.

The results of the GWR with spatial error showed a similar outcome as the GWR with spatial lag model, where the high UEQ values are located in the city down town and middle of town and the low UEQ values are located in the remote areas for the same reasons mentioned in the GWR with spatial lag model.

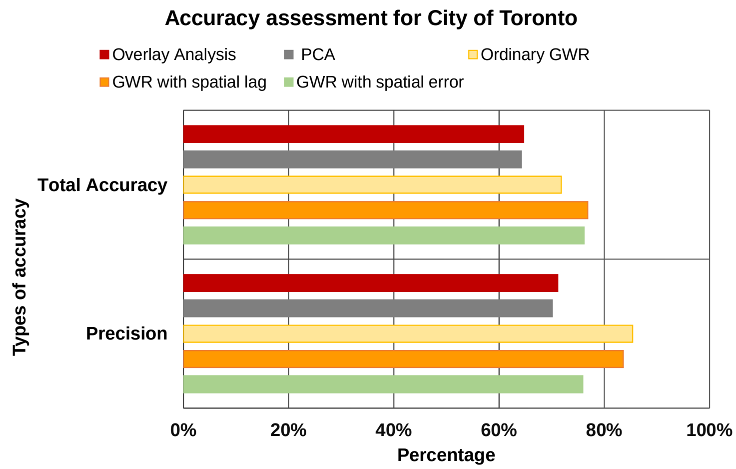


(**a**) (**b**)

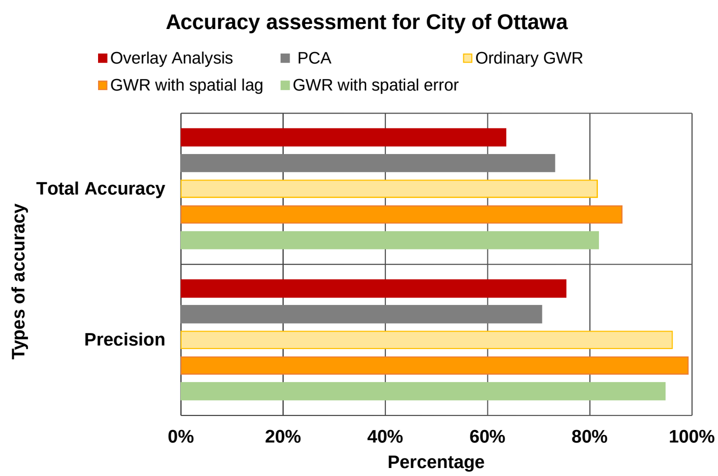
Figure 4.10: The UEQ derived using the GWR with spatial error method. (**a**) The UEQ in the City of Toronto; (**b**) the UEQ in the City of Ottawa.

4.5.4 UEQ Results Validation

As mentioned in Section 4.4.2, three socioeconomic parameters, including education level, family income and land values, were used to validate the UEQ results. The evaluation of binary classifiers approach was used to evaluate the UEQ. The results of GIS overlay, PCA and GWR (ordinary GWR, the GWR with spatial lag model and the GWR with spatial error model) were validated using socioeconomic parameters as a reference for this study. Since we are looking to highlight the higher UEQ areas, the mean values were used as a threshold to derive the higher UEQ areas. Figure 4.11 emphasizes the overall precision and accuracy of the aforementioned methods with respect to reference in this study.



(a)



(b)

Figure 4.11: The UEQ results validation. (**a**) The City of Toronto; (**b**) the City of Ottawa.

Figure 4.12 shows the reference layer and the high value of the reference layer in the two cities (the City of Toronto and the City of Ottawa). The distribution of the reference layer in the City of Toronto revealed that the highest values were found in the city centre and the west side of the city, while most of the low UEQ values were found in the northeast and northwest of the city. On the other hand, the distribution of the reference layer in the City of Ottawa revealed that the highest values were found in the city centre and the middle portions of the city, while the majority of the low UEQ values were found in the west side of the city.

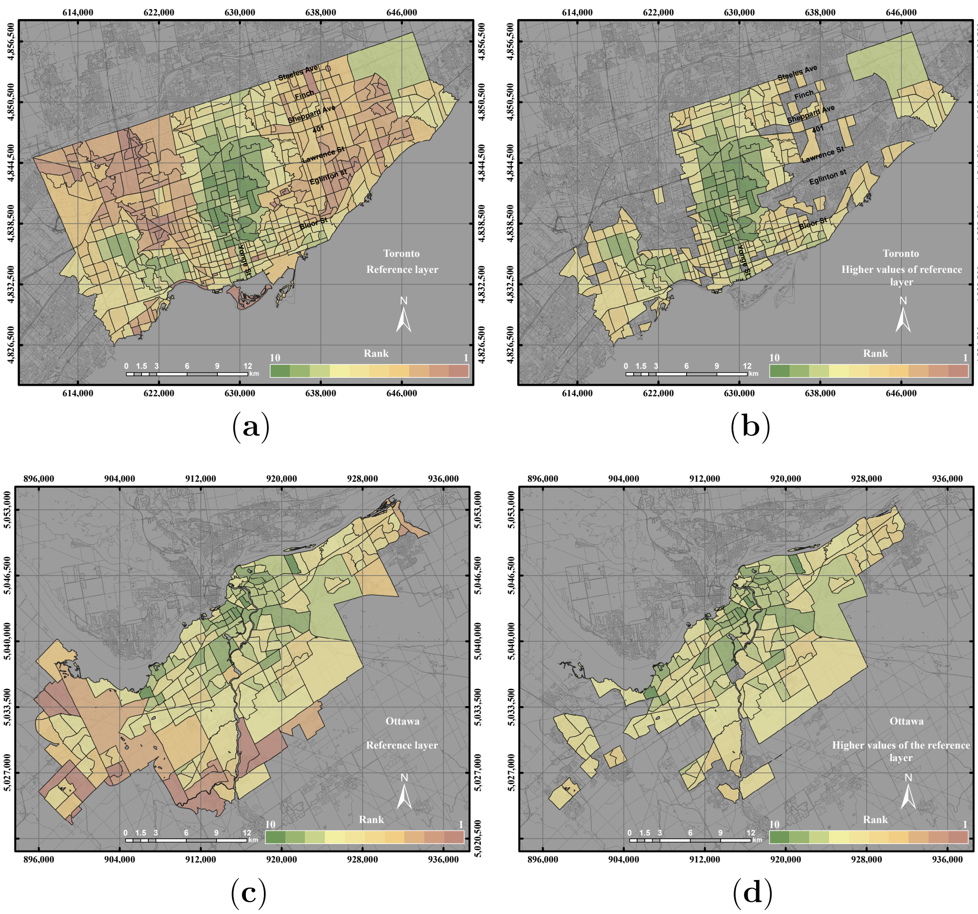
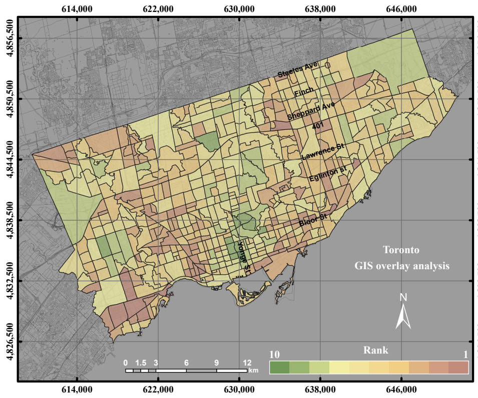
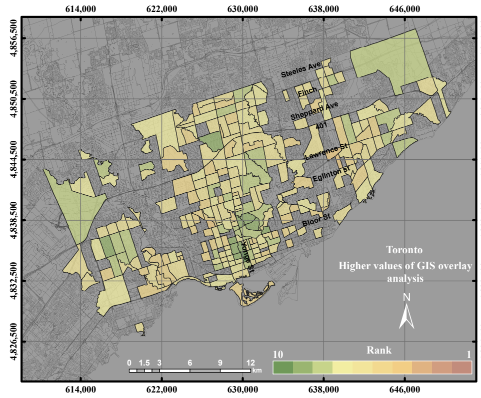


Figure 4.12: The reference layer and the higher than the mean of reference layer: (**a**) the reference layer in the City of Toronto; (**b**) the reference layer higher than the mean in the City of Toronto; (**c**) the reference layer in the City of Ottawa; (**d**) the reference layer higher than the mean in the City of Ottawa.

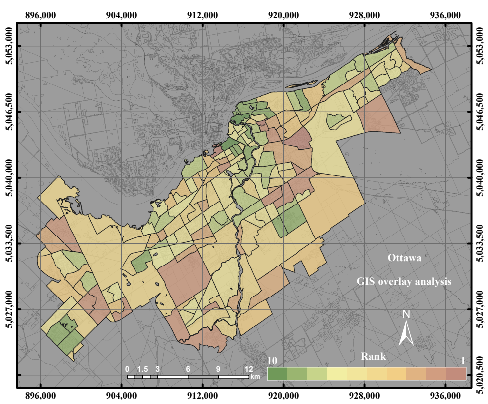
Figure 4.13 shows the GIS overlay analysis and the higher values of GIS overlay in the two cities. A few areas that have high UEQ values were located in the north and east of the city, as mentioned in Section 4.5.1. The precision and accuracy measured were found to be 71% and 65%, respectively, for the GIS overlay method in the City of Toronto. The precision and accuracy measured were found to be 75% and 63%, respectively, for the GIS overlay method in the City of Ottawa. That is mainly because that GIS overlay method considered all of the parameters, which have a negative and a positive relationship with respect to the reference layer. In addition, the parameters that have a negative relationship with respect to the reference layer might influence the overall result.



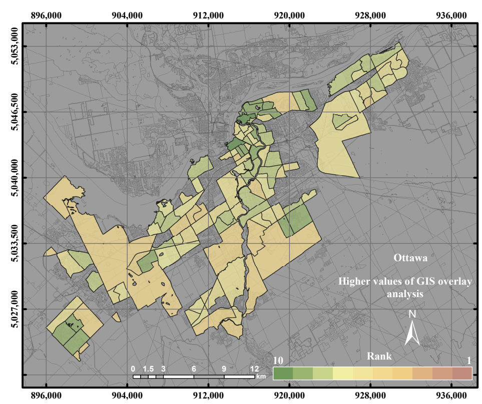
(**a**)



(**b**)



(**c**)



(**d**)

Figure 4.13: The UEQ derived using the GIS overlay method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa.

Figure 4.14(b) shows higher UEQ ranking derived using the PCA in the City of Toronto and the higher values of the PCA found in the centre, north, northwest and northeast portions of the city. The overall result of the PCA method yielded a lower precision and accuracy by 1% than the GIS overlay method and 6%–15% than GWR, GWR with spatial error and GWR with spatial lag, respectively, as shown in Figure 4.11(a). That is mainly because the PCA method does not consider 100% of the total variance. However, the rest of the methods mentioned above, including the GIS overlay method, ordinary GWR, GWR with spatial error and GWR with spatial lag, used all of the parameters. In the City of Ottawa, the PCA reported a lower precision 5% and higher accuracy by 10% than the GIS overlay method. However, PCA reported a lower precision and accuracy by 20%–25% with respect to ordinary GWR, GWR with spatial error and GWR with spatial lag for the same, as shown in Figure 4.11(b).

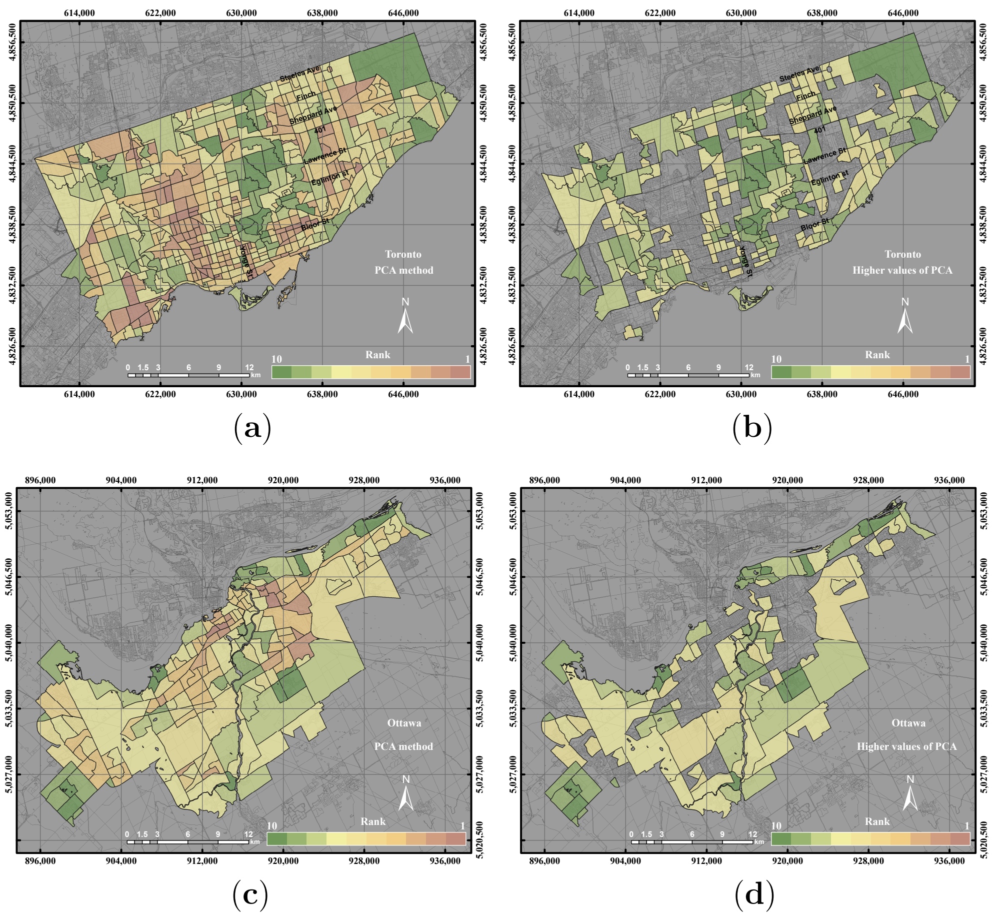


Figure 4.14: The UEQ derived using the PCA method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa.

In the City of Toronto, the ordinary GWR revealed higher precision and accuracy than the GIS overlay method and PCA method up to 14% and 7%, respectively, as shown in Figure 4.11(a). Moreover, the ordinary GWR represented higher precision up to 1% than the GWR with spatial lag model and 9% precision with respect to the GWR with spatial error model. However, the accuracy of ordinary GWR reported a lower precision up to 5% with respect to the GWR with spatial lag model and the GWR with spatial error model. That occurred by investigating the ordinary GWR and the higher values of ordinary GWR with respect to the reference layer and the higher values of the reference layer. The ordinary GWR in the City of Toronto showed that the higher values of UEQ are located in the centre, north and west of the city, as shown in Figure 4.15(b), which is visually correlated with the reference layer.

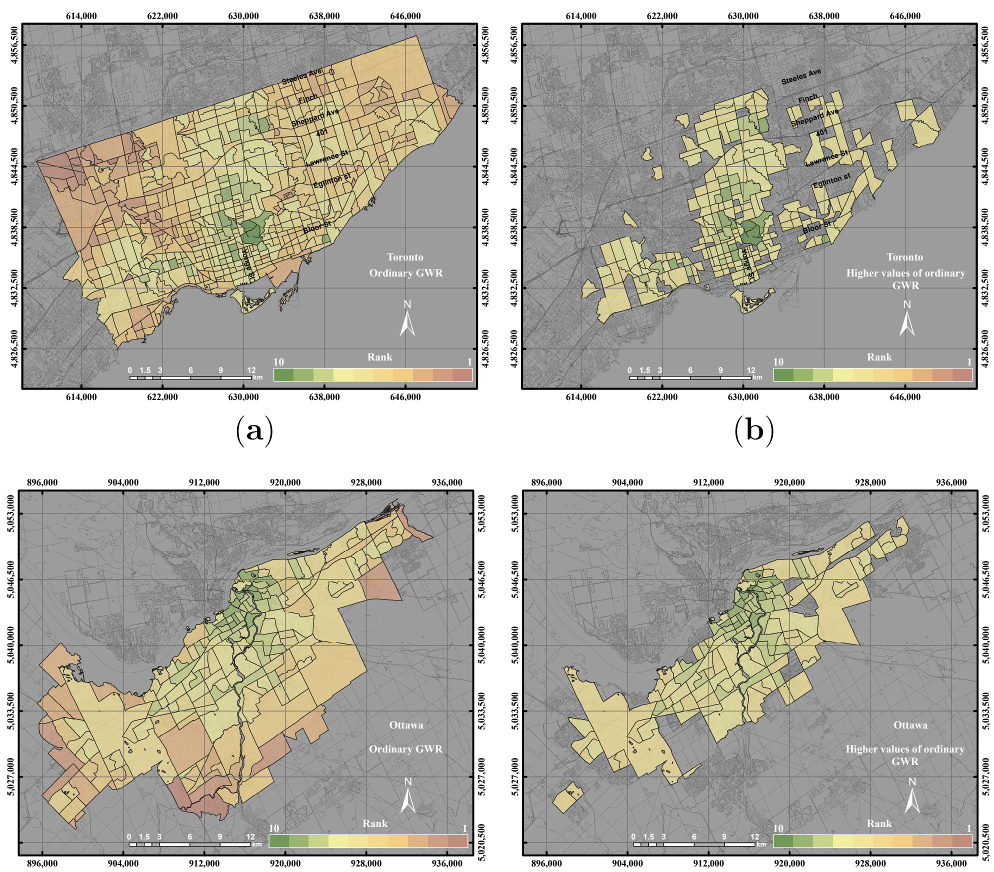


Figure 4.15: The UEQ derived using the ordinary GWR method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa.

On the other hand, in the City of Ottawa, the ordinary GWR demonstrated higher precision and accuracy than the GIS overlay method and PCA method up to 20% and 17%, respectively, as shown in Figure 4.11(b). However, the ordinary GWR revealed lower precision and accuracy up to 4% than the GWR with spatial lag model and 1% precision and accuracy with respect to the GWR with spatial error model. The ordinary GWR showed better results than GIS overlay and PCA mainly because ordinary GWR considers the spatial weight component in the method.

Figure 4.16 a and b shows the GWR with spatial lag model and the higher values of GWR with spatial lag model in the City of Toronto.

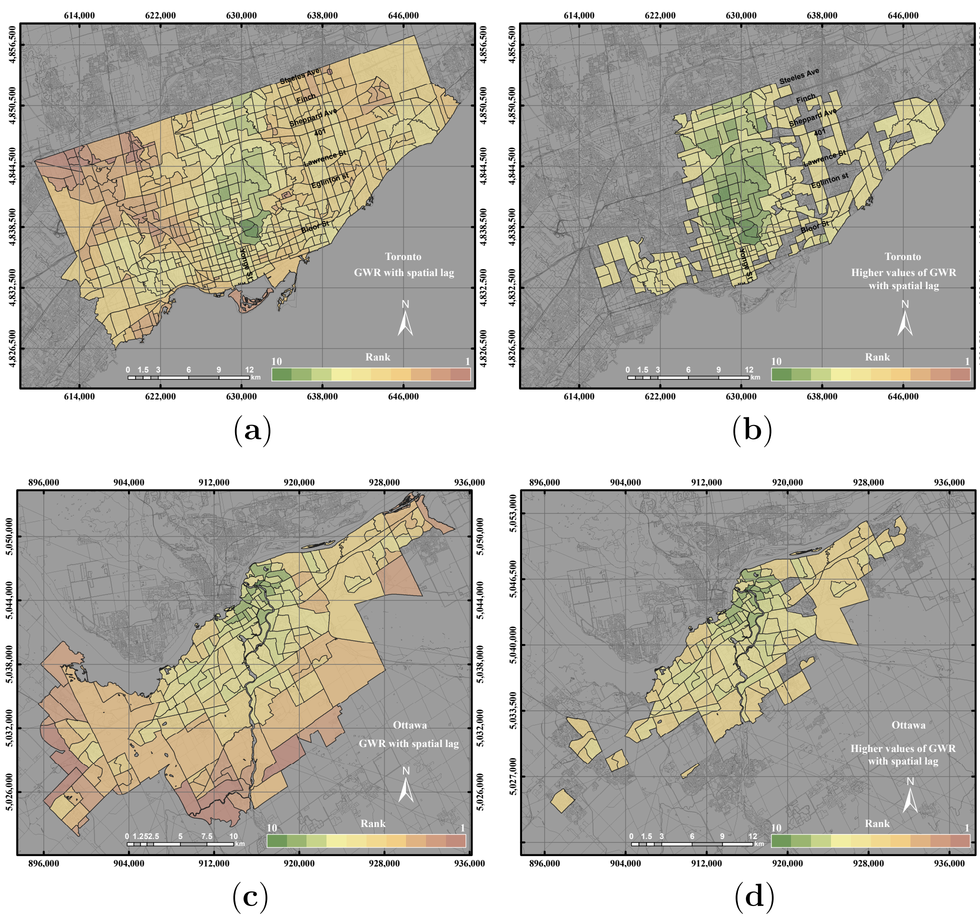


Figure 4.16: The UEQ derived using the GWR with spatial lag method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa.

As shown in Figure 4.16(b) the higher UEQ values are located in the centre, north and west of the city, which is also visually correlated with the reference layer. Thus, the precision and accuracy of the GWR with spatial lag model reported better results than GIS overlay and PCA by 15% and 8%, respectively, and 1%–5% with respect to GWR with spatial error and ordinary GWR, respectively. That is mainly because the GWR with spatial lag model adjusts the spatial heterogeneity by including an autocorrelation coefficient as mentioned previously in Geographically-Weighted Regression with Spatial Lag Model Section. The results of the City of Ottawa, on the other hand, yielded higher UEQ values that are located in the centre and the middle of the city, as the ordinary GWR. The precision and accuracy of the GWR with spatial lag model reported better results than GIS overlay and PCA by 15% and 20%, respectively, and 5% with respect to both GWR with spatial error and ordinary GWR.

The precision and accuracy of the GWR with spatial error model both revealed 76% in the City of Toronto, but 94% and 81%, respectively, in the City of Ottawa, as shown in Figure 4.11. The higher values of the GWR with spatial error model in the City of Ottawa were located in the centre and the middle of the city, the same as the ordinary GWR and the GWR with spatial lag model, as shown in Figure 4.17(d). On the other hand, the higher values of the GWR with spatial error model in the City of Toronto emerged along Yonge Street, as shown in Figure 4.17(b). Figure 4.11 shows that the GWR with spatial error model revealed better precision and accuracy than GIS overlay and PCA. However, the GWR with spatial error model represents lower precision and accuracy with respect to the GWR with spatial lag model.

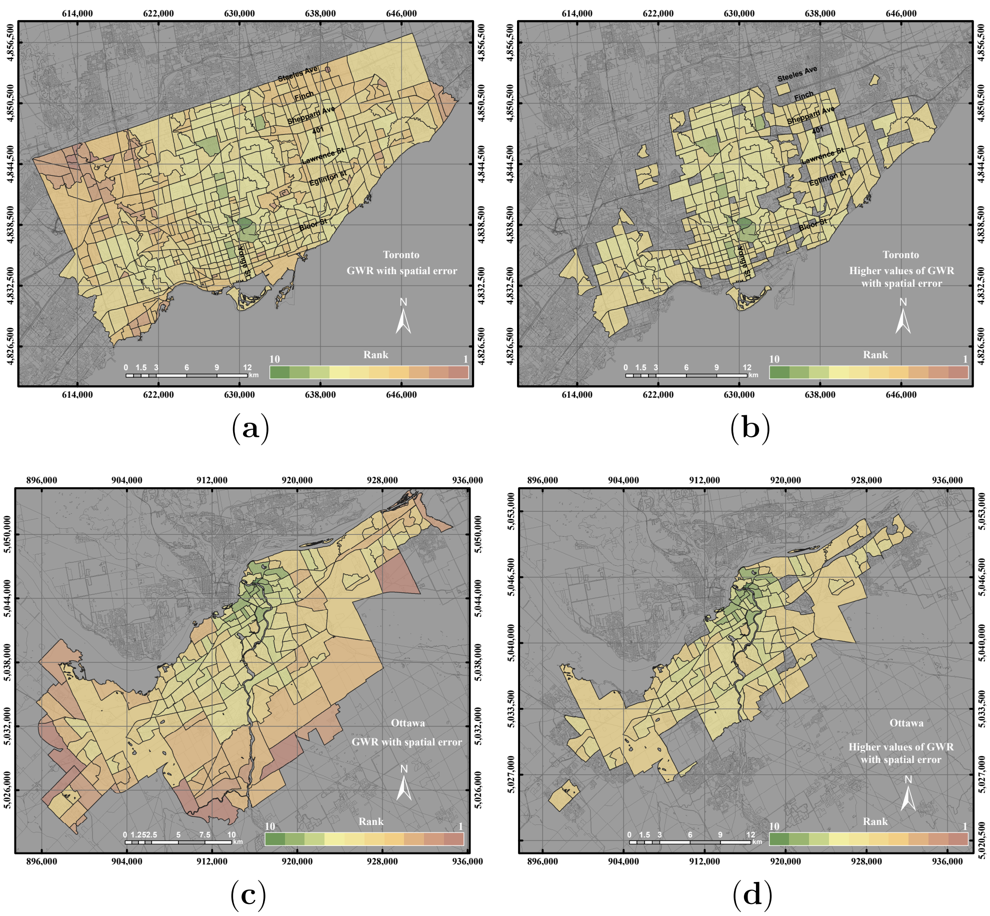


Figure 4.17: The UEQ derived using the GWR with spatial error method: (**a**) the derived UEQ in the City of Toronto; (**b**) UEQ zones higher than the mean in the City of Toronto; (**c**) the derived UEQ in the City of Ottawa; (**d**) UEQ zones higher than the mean in the City of Ottawa.

Besides the successful attempted methods used in this research work, there are several potential draw backs: (1) the lack of data is always an issue that may influence the final results; (2) census socioeconomic data are usually related to administrative units and can be changed in a short period, which makes it difficult to have them available worldwide; (3) remote sensing, GIS and socioeconomic data need data transformation from raster to vector or from vector to raster, which could cause an individual loss of spatial information; (4) the distance-based weighted algorithm is more applicable to a flat surface, so all of the polygons need to be projected in advance for the output to be correct;

(5) the authors have previously investigated the use of linear and nonlinear regression to run the relationship between the derived UEQ with respect to the socio-economic data (reference data). However, there is no meaningful trend found in the two cities that thus reveals the inappropriate use of a linear or non-linear model in this particular case study.

4.6 Conclusions

This paper epitomizes the use of the GIS overlay, PCA and GWR techniques to assess UEQ with two case studies in Ontario, Canada. The main contribution of this research work is to investigate a new method to normalize various data derived from remote sensing, GIS and census data. New approaches of GWR techniques, including the GWR with spatial lag model and the GWR with spatial error model, were tested to assess the UEQ. The new approach was evaluated to validate the final outcomes derived from the above-mentioned methods. GWR is an intellectual framework that considers the spatial relationship among the polygons in each parameter. The GWR with spatial lag model was mainly used to provide homogeneous results by incorporating the spatial lag of the dependent variable into the GWR. Therefore, the GWR with spatial lag model is capable of producing better outcomes than other unweighted integration techniques.

The GWR with spatial error model was used in this study to correct the spatial autocorrelation in the spatial data. It was found that the middle of town, north of town and southwest areas have high UEQ in the City of Toronto. However, higher UEQ was found in the city centre and middle of town within the City of Ottawa. The results illustrated that the GWR with spatial lag model significantly improved the final outcomes with respect to unweighted methods, including GIS overlay and PCA up to 15% (precision) and 8% (accuracy) in the City of Toronto and 15% (precision) and 20% (accuracy) in the City of Ottawa. Moreover, the GWR with spatial lag model also improved the final outcomes with respect to weighted methods, including ordinary GWR and GWR with spatial error model up to 1% (precision) to 5% (accuracy) in the City of Toronto and 5% (precision and accuracy) in the City of Ottawa. Thus, the GWR with spatial lag model can be used to integrate multiple parameters for UEQ purposes more accurately than the unweighted integration techniques.

Besides the success of the attempted methods used in this research work, there are several potential draw backs: (1) the lack of data is always an issue that may influence the final results; (2) census socioeconomic data are usually related to administrative units and can be changed in a short period, which makes it difficult to have them available worldwide; (3) remote sensing, GIS and socioeconomic data need data transformation from raster to vector or from vector to raster, which could cause an individual loss of spatial information; (4) the distance-based weighted algorithm is more applicable to a flat surface, so all of the polygons need to be projected in advance for the output to be correct; (5) the authors have previously investigated the use of linear and nonlinear regression to run the relationship between the derived UEQ with respect to the socio-economic data (reference data). However, there is no meaningful trend found in the two cities that thus reveals the inappropriate use of a linear or non-linear model in this particular case study.

Municipalities and decision makers can consider the proposed approach to derive the UEQ for sustainable planning in many countries. However, there is always a need for new improvement to derive better precision and accuracy in the future. Therefore, updated remote sensing and GIS data are important for better results; also, integration between weighted and GWR can be a promising method to enhance the final outcomes of UEQ; future work can be focused on modelling UEQ for an arid or cold region environment/country since there are some parameters that may not be applicable in those areas. In conclusion, remote sensing and GIS techniques are useful tools to model UEQ. Spatial weighting methods further can enhance the capability to estimate UEQ in a more accurate manner.

Chapter 5

Conclusions and Future Work

This chapter concludes the findings of this research. In addition, further recommendations are proposed in this section for future research.

5.1 Conclusions

As mentioned in chapter one, there is a lack of researches that provided an in-depth discussion on the UEQ parameters , which can be used to assess the UEQ. The majority of the scholars mainly utilized Principal Component Analysis (PCA), Geographic Information System (GIS) analysis or Multi-Criteria Evaluation (MCE) techniques to integrate UEQ parameters (Nichol and Wong, 2009; Fobil *et al.*, 2011; Lo, 1996; Rinner, 2007), where there exist certain limitations for all these integration techniques. 1) PCA itself produces unweighted components, which may not highlight those important parameters; 2) PCA does not work properly in nonlinear relationships; and finally, 3) the minimum number of components is indeterminable (Faisal and Shaker, 2017). GIS overlay method does not consider correlation among parameters nor give weight to the parameters. MCE mainly relies on the user's input of the attribute values of the parameters. Normalizing the parameters is indeed essential to standardize the observations of each parameter. Several research work including (Fobil *et al.*, 2011; Rinner, 2007; Moore *et al.*, 2006; Lo, 1996; Liang and Weng, 2011) did not highlight the importance or consider the data normalization. The result validation of the UEQ studies have not been fully investigated. Though a few existing research addressed result validation (Nichol and Wong, 2009; Rahman *et al.*,

2011) using e-mail questionnaire or field-based questionnaire. However, most of the UEQ studies (Fobil *et al.*, 2011; Rinner, 2007; Moore *et al.*, 2006; Lo, 1996; Liang and Weng, 2011) did not perform any field survey or even result validation. That is mainly because collecting field data is time consuming and budget dependent. Furthermore, these field data can be inaccurate to test the outcomes of UEQ if the data samples being collected are not representative, which may lead to bias results. In this regard, several goals were proposed in this dissertation in order to fill the current gaps found in previous studies. 1) To examine the relationship of multiple UEQ parameters derived from remote sensing, GIS and socio-economic data; 2) to evaluate some of the existing methods (e.g. linear regression, GIS overlay and PCA) for assessing and integrating multiple UEQ parameters; 3) to propose a new method to weight urban and environmental parameters obtained from different data sources; 4) to develop a new method to validate urban and environmental parameters with socio-economic indicators for UEQ assessment in two cities in Ontario, Canada. The methodological contribution of this research work can be summarized as the following:

- **A new method was designed to highlight the industrial agglomeration within urban areas and to estimate the socio-economic data (i.e., real GDP, total population and total employment) based on remote sensing data.** Built-up index parameter was investigated to identify the industrial areas using remote sensing and GIS techniques. The results showed that built-up index can be investigated to compute the socio-economic data (i.e., real GDP, total population and total employment) in Canada and these results showed promising coefficient regression up to 0.83 (R^2) in most of the large cities in Canada. However, some cities including Calgary and Edmonton have a rapid expansion of gross income from the oil mining industry that does not require a large piece of land for manufacturing within the city. Therefore, the built-up index cannot adequately represent the aforementioned socio-economic data in cities that mainly depend on oil mining industry. Taken together, our results can be used as a general indication for the federal/municipal authorities, which are aiming at or targeting a specific real GDP with respect to the planned industrial areas for city management.

- **Evaluate the urban, environmental parameters and socio-economic indicators obtained from a different source of data.** Several remote sensing and GIS data were explored in order to fully understand the concept of UEQ. As mentioned in chapter 1, several scholars have discussed the optimal urban and environmental parameters for the UEQ assessment (Zavadskas *et al.*, 2007; Gabrielsen and Bosch, 2003; Indicators for Sustainable Cities, 2015). Urban planners and policymakers, assigned a sheer number of indicator frameworks, which vary in their approach for measuring UEQ and their selection of indicators (Zavadskas *et al.*, 2007). Most of the indicator frameworks are valid and representative for UEQ. However, some of the indicator frameworks are built for a particular location (Berrini and Lorenzo, 2010; Dekker *et al.*, 2003; Indicators for Sustainable Cities, 2015). Since the urban and environmental parameters and socio-economic indicators were obtained in various scale level, data normalization is needed to evaluate the significance of each parameter.

- **A new approach was examined to normalize the data derived from remote sensing and GIS data using Z-score model and linear interpolation technique.** The primary goal of selecting UEQ indicators is to assess the performance of these indicators. Therefore, the indicators are needed to be standardized and addressed so that they are presented on the same scale. In this manner, the selected indicators can be validated and enhanced to serve UEQ (Yigitcanla and Lönnqvist, 2013). Moreover, standardization can also help understanding of the indicators (Pires *et al.*, 2014). In this dissertation, the normalized data were integrated using GIS overlay, PCA and weighted methods including ordinary GWR, GWR with spatial lag model and the GWR with spatial error model to compute the UEQ outcomes.

- **A new approach was designed to validate the final results that calculated from the above mentioned methods. Socio-economic indicators, including family income, the degree of education and land value, were used as a reference to validate the outcomes derived from the five integration methods.** Several existing researchers found that socio-economic indicators including education level and income are required for UEQ assessment (Adelle and Pallemaerts, 2009). People with more education and income are more likely to support high quality environment (Kahn and Matsusaka, 1997; Kahn, 2002).

For example, richer urbanites are more likely to support high quality urban areas and purchase good cars that produce less pollution per kilometer (Kahn, 2007). Consequently, several studies showed that income indicator has high relationship up to 0.91 with GDP, car and house ownership in 158 nations in 1996 (Kahn, 2007; Kahn and Matsusaka, 1997; Kahn, 2002). Education provides the tools for people to access and understand information about how environmental hazards affect their wellbeing. As a result, rising educational level can help increase the awareness of individuals for better quality regions (Becker and Mulligan, 1997). Studies also observed a high correlation between the level of education and voting, since people with high education are more likely aware of public/political issues that may influence their environment quality (Kahn, 2002). Thus, socio-economic indicators are essential to assess the UEQ for any urban areas. The finding of this research work showed that weighted methods represented up to 20% better results than unweighted methods and GIS overlay approach showed better outcomes than PCA up to 4% (precision) and 2% (accuracy) in the City of Toronto and City of Ottawa. These finding suggested that GIS overlay can be presented as a better method than PCA concerning the integration of multiple parameters. Moreover, geographically weighted methods are deemed to be a better method with respect to weighted methods including GIS overlay and PCA.

- **New methods were proposed to weight urban and environmental parameters that obtained from a different source of data.** GWR techniques, including the ordinary GWR, GWR with spatial lag model and the GWR with spatial error model were tested to assess the UEQ in the City of Toronto and City of Ottawa. The observed results showed that GWR with spatial lag model significantly improved the final outcomes with respect to unweighted methods, including GIS overlay and PCA for up to 15% (precision) and 8% (accuracy) in the City of Toronto and 15% (precision) and 20% (accuracy) in the City of Ottawa. Moreover, the GWR with spatial lag model also improved the final outcomes with respect to weighted methods, including ordinary GWR and GWR with spatial error model for up to 1% (precision) to 5% (accuracy) in the City of Toronto and 5% (precision and accuracy) in the City of Ottawa. Thus, the GWR with spatial lag model can be used to integrate multiple parameters for UEQ assessment more accurately than the unweighted integration techniques. In conclusion, municipalities and decision makers can consider the proposed approach to derive the UEQ for sustainable planning in many countries. Remote sensing and GIS techniques are useful tools to model UEQ. Spatial weighting methods further can enhance the capability to estimate UEQ in a more accurate manner.

In conclusion, this research work broached the term UEQ for urban planners and decision making. The research work exuberance scientific methods that investigate data acquisition, data processing, leading to the final outcome. Choosing the parameters or indicators is generally not relied on the individual or the researcher. The sets of indicators are shared among international stakeholders and institutions (Delsante, 2016). Previously several international agencies such as Environmental Sustainability Index (ESI), Environmental Performance Index (EPI), European Green Cities Index, China Urban Sustainability Index and Global City Indicator have progressively developed different set of indicator frameworks that serve different location worldwide (Indicators for Sustainable Cities, 2015). In this research work, the proposed set of parameters for urban quality evaluation allows for comparison between the various locations under specific circumstances. The investigated study areas should be chosen based on specific local conditions (Delsante, 2016).

The final outcomes showed that the suggested methods can be an essential tool for concisely assessing urban environment quality in Canada. The proposed methods can be beneficial regarding sustainability for urban planners and decision making by evaluating performances of UEQ for each metropolitan regions and adopting the best political actions. The higher UEQ areas represent the good locations within the urban areas that can be used as a reference to promote all the other low UEQ regions. The research work, also highlighted that remote sensing technique could be utilized as a tool to derive some of the socio-economic indicators such as GDP, population and employment rate as representing in chapter 2. These socio-economic indicators including GDP, population and employment rate also showed high correlation with respect to income (Kahn, 2007; Kahn and Matsusaka, 1997; Kahn, 2002), which was generated as one of the indicators to validate the final UEQ outcomes in this research work. In context, the UEQ assessment is an intellectual tool for the disciplines of architecture and urban design. If the UEQ evaluation is monitored over time and/or compared with other study areas, the results can help public authorities and other relevant stakeholders for proper actions.

5.2 Limitations and Future Work

Besides the success of the attempted methods used in this research work, there are few limitations in regards to the research study that needed to be considered in the near future:

1. Water bodies and bare soil all have high built-up index values that may cause confusion with the impervious surfaces. If the method presented in Chapter 2 is applied elsewhere and no GIS data exists, it would likely cause problems, as water bodies or bare soil could be classified as built-up areas, and the relationship with GDP would be affected.

2. The lack of data is always an issue that may influence the final results. Data availability is another significant issue that needs to be considered when indicators are selected for UEQ assessment. Pires *et al.* (2014) highlighted that unavailable data sources could cause a biased or unreliable estimate for UEQ. Researchers consensus that indicator sets need to be locally relevant to the city or municipality (Campbell, 1999; Camagni, 2002). Therefore, the indicators should be clear, understandable and obtainable at a reasonable cost-benefit ratio and must be able to reflect every aspect of urban development (Mega and Pedersen, 1998).

3. Census socioeconomic data are usually related to administrative units and can be changed in a short period, which makes it difficult to have them available worldwide. Some scholars emphasized that indicators with extensive political support were more successful than those proposed by academic institutions or non-government agencies (Hiremath *et al.*, 2013).

4. Remote sensing, GIS and socioeconomic data need data transformation from raster to vector or from vector to raster, which could cause an individual loss of spatial information. Most of the current vectorization software are time consuming, budget dependence and some time may steered to unreliable results. That is mainly because of the current vectorization software that are semi-automatic and predominantly require an intervention from human (Taie *et al.*, 2011).

5. The distance-based weighted algorithm is more applicable to a flat surface, so all of the polygons need to be projected in advance for the output to be correct. Distance-based weight method may not be applicable for unprojected geospatial datasets mainly because they are located on the three-dimensional geodetic coordinate system. In addition, some study areas such as Canada or USA are too wide, and the geospatial datasets are distributed within a wide range from the East Coast to the West Coast. In that case, the line distance method may not also be applicable due to the curvature of the earth (Luc and Sergio, 2014).

6. The authors have previously investigated the use of linear and nonlinear regression to run the relationship between the derived UEQ with respect to the socio-economic data (reference data). However, there is no meaningful trend found in the two cities (City of Toronto and City of Ottawa) that thus reveals the inappropriate use of a linear or non-linear model in this particular case study.

To further improve the findings of this research work, future research may be required to be studied in the following areas:

1. Updated remote sensing and GIS data are essential for better results. In addition, several parameters can be further investigated to help authorities and decision making derived more trustworthy UEQ results. Plans of flood risk, traffic volumes and pollution levels, which include water and air pollution are very important to improve the overall UEQ outcomes.

2. Integration between weighted and GWR can be a promising method to enhance the final outcomes of UEQ.

3. Other regression analyses (such as nonlinear regression) can be explored depending on the nature of study area and the socio-economic indicators being studied.

4. Other statistical algorithm, such as stepwise regression can be utilized to narrow down the variety of parameters

5. Logistic regression can be further investigated for the binary classification of each parameter to compute the significant of each observation.

6. Future work can be focused on modeling UEQ for an arid or cold region environment/country since some parameters being used in this dissertation may not be applicable to those areas. Moreover, some countries/ cities does not have proper GIS data. Therefore, remote sensing can be an essential technique to derive some of the UEQ parameters.

References

Adelle, C. and Pallemaerts, M., 2009. Sustainable Development Indicators. An Overview of Relevant Framework Programme Funded Research and Identification of Further Needs in View of EU and International Activities. *European Communities: European Commission*.

Anand, P., *et al.*, 2008. The measurement of capabilities, In Arguments for a Better World. Essays in Honor of Amartya Sen. Oxford University Press: Oxford, UK.

Anselin, L. and Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, 155, 237–290.

Luc, A. and Sergio, R., 2014. Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL. GeoDa Press LLC: Chicago, IL, USA.

Barsi, J.A., *et al.*, 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments. *In*: *Optics & Photonics 2005*, 58820E–58820E.

Becker, G. and Mulligan, C., 1997. The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3), 729–758.

Bederman, S.H. and Hartshorn, T.A., 1984. Quality of life in Georgia: the 1980 experience. *Southeastern Geographer*, 24 (2), 78–98.

Berrini, M. and Lorenzo, B., 2010. Measuring urban sustainability: Analysis of the European Green Capital Award 2010 and 2011 application round. *Ambiente Italia* 1–44.

Bhatti, S.S. and Tripathi, N.K., 2014. Built-up area extraction using Landsat 8 OLI imagery. *GIScience & Remote Sensing*, 51 (4), 445–467.

Borst, R. and McCluskey, W., 2007. Comparative evaluation of the comparable sales method with geostatistical valuation models. *Pacific Rim Property Research Journal*, 13 (1), 106–129.

Camagni, R., 2002. On the concept of territorial competitiveness: sound or misleading?. *Urban studies*, 39 (13), 2395–2411.

Campbell, S., 1996. Green cities, growing cities, just cities?: Urban planning and the contradictions of sustainable development.*Journal of the American Planning Association*, 62 (3), 296–312.

Canadian Centre for Energy Information, 2012. Evolution of Canada's Oil and Gas Industry. [http://www.capp .ca/getdoc.aspx?dt=PDF&docID=206748] [Online; accessed June 14, 2014].

Chander, G., Markham, B.L. and Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113 (5), 893–903.

Charlton, M., Fotheringham, S. and Brunsdon, C., 2009. Geographically weighted regression. *White paper. National Centre for Geocomputation. National University of Ireland Maynooth*.

Cheadle, C., *et al.*, 2003. Analysis of microarray data using Z score transformation. *The Journal of molecular diagnostics*, 5 (2), 73–81.

Chen, X.L., *et al.*, 2006. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote sensing of environment*, 104 (2), 133–146.

Cihlar, J., 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, 21 (6-7), 1093–1114.

Cliff, N., 1988. The eigenvalues-greater-than-one rule and the reliability of components. *Psychological bulletin*, 103 (2), 276.

Coll, C., *et al.*, 2010. Validation of Landsat-7/ETM+ thermal-band calibration and atmospheric correction with ground-based measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 48 (1), 547–555.

Cullen, J.B. and Levitt, S.D., 1999. Crime, urban flight, and the consequences for cities. *Review of economics and statistics*, 81 (2), 159–169.

Curran, P. and Steven, M., 1983. Multispectral remote sensing for the estimation of green leaf area index [and discussion]. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 309 (1508), 257–270.

Czajkowski, K.P., *et al.*, 2004. Estimating environmental variables using thermal remote sensing. *Thermal remote sensing in land surface processes*, 11–32.

Dekker, S., Jacob, J., Klassen, E., Miller, H., Thielen, S. and Their, W.W., 2012. Indicators for Sustainability. [http://sustainablecities.net/wp-content/uploads/2015/10/indicators-for-sustainability-intl-case-studies-final.pdf] [Online; accessed June 11, 2017].

Delsante, I., 2016. Urban environment quality assessment using a methodology and set of indicators for medium-density neighbourhoods: a comparative case study of Lodi and Genoa. *Ambiente Construído*, 16 (3), 7–22.

De Smith, M.J., 2004. Distance transforms as a new tool in spatial analysis, urban planning, and GIS. *Environment and planning B: Planning and design*, 31 (1), 85–104.

Diener, E. and Suh, E., 1997. Measuring quality of life: Economic, social, and subjective indicators. *Social indicators research*, 40 (1), 189–216.

Din, A., Hoesli, M. and Bender, A., 2001. Environmental variables and real estate prices. *Urban studies*, 38 (11), 1989–2000.

Din Özdemir, A., 2007. Urban sustainability and open space networks. *Journal of Applied Sciences*, 7 (23), 3713–3720.

Faisal, K. and Shaker, A., 2014. The use of remote sensing technique to predict Gross Domestic Product (GDP): An analysis of built-up index and GDP in nine major cities in Canada. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40 (7), 85.

Faisal, K. and Shaker, A., 2017. An Investigation of GIS Overlay and PCA Techniques for Urban Environmental Quality Assessment: A Case Study in Toronto, Ontario, Canada. *Sustainability*, 9 (3), 380.

Faisal, K., Shaker, A. and Habbani, S., 2016. Modeling the Relationship between the Gross Domestic Product and Built-Up Area Using Remote Sensing and GIS Data: A Case Study of Seven Major Cities in Canada. *ISPRS International Journal of Geo-Information*, 5 (3), 23.

Fobil, J.N., *et al.*, 2011. Neighborhood urban environmental quality conditions are likely to drive malaria and diarrhea mortality in Accra, Ghana. *Journal of environmental and public health*, 2011.

Fotheringham, A.S., Brunsdon, C. and Charlton, M., 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Fung, T. and Siu, W., 2000. Environmental quality and its changes, an analysis using NDVI. *International Journal of Remote Sensing*, 21 (5), 1011–1024.

Fung, T. and Siu, W.L., 2001. A study of green space and its changes in Hong Kong using NDVI. *Geographical and Environmental Modelling*, 5 (2), 111–122.

Gabrielsen, P. and Bosch, P., 2003. Environmental indicators: typology and use in reporting. *EEA, Copenhagen*.

Ghosh, T. and Elvidge, C.D., 2010. Estimating the information and technology development index (IDI) using nighttime satellite imagery. *Proceedings of the Asia-Pacific Advanced Network*, 30, 143–171.

Giannini, M., *et al.*, 2015. Land Surface Temperature from Landsat 5 TM images: Comparison of different methods using airborne thermal data. *J. Eng. Sci. Technol. Rev*, 8, 83–90.

Green, N.E., 1957. *Aerial photographic interpretation and the social structure of the city*. American Society of Photogrammetry.

Guesgen, H.W. and Albrecht, J., 2000. Imprecise reasoning in geographic information systems. *Fuzzy Sets and Systems*, 113 (1), 121–131.

Hansen, M.C. and Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment*, 122, 66–74.

Hardisky, M., Klemas, V. and Smart, R.M., 1983. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina alterniflora canopies. *Photogrammetric Engineering and Remote Sensing*, 49, 77–83.

Harmon, J.E. and Anderson, S.J., 2003. *The design and implementation of geographic information systems*. John Wiley & Sons.

Harris Consulting, 2007. City of Victoria: Downtown and Downtown Fringe Retail Commercial and Hotel Demand Forecast. [http://www.victoria.ca/cityhall/pdfs/plnpln_downtown_anlyss_appc2.pdf] [Online; accessed June 25, 2014].

He, C., *et al.*, 2010. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. *Remote Sensing Letters*, 1 (4), 213–221.

Hiremath, R., Balachandra, P., Kumar, B., Bansode, S. and Murali, J., 2014. Indicator-based urban sustainability-A review. *Energy for sustainable development*, 17(6), 555–563.

Huang, W., Zeng, Y. and Li, S., 2015. An analysis of urban expansion and its associated thermal characteristics using Landsat imagery. *Geocarto International*, 30 (1), 93–103.

In-Depth Report: Indicators for Sustainable Cities, 2015. Science for Environment Policy.[http://ec.europa.eu/environment/integration/research/newsalert/pdf/indicators-for-sustainable-cities-IR12-en.pdf] [Online; accessed June 15, 2017].

Initiative, O.B.L., 2011. compendium of OECD well-being indicators. *Paris: Organisation for Economic Co-operation and Development*.

Irvine, K.N., *et al.*, 2009. Green space, soundscape and urban sustainability: an interdisciplinary, empirical study. *Local Environment*, 14 (2), 155–172.

Jensen, J.R., 2005. Introductory digital image processing 3rd edition. *Upper saddle river: Prentice hall*.

Kahn, M., 2002. Demographic change and the demand for environmental regulation*Journal of Policy Analysis and Management*, 21(1), 45–62.

Kahn, M.E., 2007. *Green cities: urban growth and the environment*. Brookings Institution Press.

Kahn, M. and Matsusaka, J., 1997. Demand for Environmental Goods: Evidence from Voting Patterns on California Initiatives 1 *The Journal of Law and Economics*, 40(1), 137–174.

Kahn, M.E. and Mills, E.S., 2006. *Green cities: urban growth and the environment*. Cambridge Univ Press.

Kjaersgaard, J. and Allen, R., 2009. Loading the Landsat thermal band preprocessed using the EROS LPGS system. *Kimberly R&E Center, University of Idaho. Report*.

Krause, E., 1987. Taxicab geometry: An adventure in non-Euclidean geometry, DoverPublications. com. .

Landefeld, S.J., Seskin, E.P. and Fraumeni, B.M., 2008. Taking the pulse of the economy: Measuring GDP. *The Journal of Economic Perspectives*, 22 (2), 193–193.

Landorf, C., Brewer, G. and Sheppard, L.A., 2008. The urban environment and sustainable ageing: critical issues and assessment indicators. *Local Environment*, 13 (6), 497–514.

Lawrence, R.L. and Ripple, W.J., 1998. Comparisons among vegetation indices and bandwise regression in a highly disturbed, heterogeneous landscape: Mount St. Helens, Washington. *Remote Sensing of Environment*, 64 (1), 91–102.

Leask, A. and Fyall, A., 2006. *Managing world heritage sites*. Routledge.

Leidelmijer, K., Van Kamp, I. and Marsman, G., 2002. Leefbaarheid Naar een begrippenkader en Conceptuele inkadering (RIGO, RIVM).

Leymarie, F. and Levine, M.D., 1992. Fast raster scan distance propagation on the discrete rectangular lattice. *CVGIP: Image Understanding*, 55 (1), 84–94.

Li, G. and Weng, Q., 2007. Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. *International Journal of Remote Sensing*, 28 (2), 249–267.

Li, Z.L., *et al.*, 2013. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131, 14–37.

Liang, B. and Weng, Q., 2011. Assessing urban environmental quality change of Indianapolis, United States, by the remote sensing and GIS integration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4 (1), 43–55.

Liu, Q., Sutton, P.C. and Elvidge, C.D., 2011. Relationships between nighttime imagery and population density for Hong Kong. *Proceedings of the Asia-Pacific Advanced Network*, 31, 79–90.

Lo, C.P., Quattrochi, D.A. and Luvall, J.C., 1997. Application of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *International Journal of Remote Sensing*, 18 (2), 287–304.

Lo, C., 1996. Integration of Landsat Thematic Mapper (TM) Data and US Census Data for Quality of Life Assessment. *International Archives of Photogrammetry and Remote Sensing*, 31, 431–436.

Lu, B., *et al.*, 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, 28 (4), 660–681.

Lynch, A., Andreason, S., Eisenman, T., Robinson, J., Steif, K. and Birch, E., 2011. Sustainable Urban Development Indicators for the United States: Report to the Office of International and Philanthropic Innovation, Office of Policy

Development and Research, US Department of Housing and Urban Development. [http://penniur.upenn.edu/uploads/media/sustainable-urban-development-indicators-for-the-united-states.pdf] [Online; accessed June 10, 2017].

Ma, Y. and Xu, R., 2010. Remote sensing monitoring and driving force analysis of urban expansion in Guangzhou City, China. *Habitat International*, 34 (2), 228–235.

Martel, L., 2012. *The Canadian Population in 2011, Population Counts and Growth: Population and Dwelling Counts, 2011 Census*. Statistics Canada.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17 (7), 1425–1432.

Mega, V. and Pedersen, J., 1998. Urban Sustainability Indicators. *Office for Official Publications of the European Communities. Luxembourg.*

Metropolitan Housing Outlook, 2013. In-Depth Housing Analysis for Canada, the Provinces, and Nine Metropolitan Areas. [http://www.genworth.ca/en/pdfs/Metropolitan_Housing_Outlook_Autumn13_EN.pdf] [Online; accessed June 19, 2014].

Mira, R.G., *et al.*, 2005. *Housing, space and quality of life.* .

Moore, G., *et al.*, 2006. Urban environmental quality: perceptions and measures in three UK cities. *WIT Transactions on Ecology and the Environment*, 93.

Moran, M.S., *et al.*, 1992. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, 41 (2-3), 169–184.

Nakaguchi, T., 1999. The development of natural environmental indices to set targets for a local environmental plan. *Geographical Review of Japan, Series A*, 72 (2), 93–115.

Newman, P. and Kenworthy, J., 1999. *Sustainability and cities: overcoming automobile dependence*. Island Press.

Nichol, J. and Lee, C., 2005. Urban vegetation monitoring in Hong Kong using high resolution multispectral images. *International Journal of Remote Sensing*, 26 (5), 903–918.

Nichol, J. and Wong, M.S., 2009. Mapping urban environmental quality using satellite data and multiple parameters. *Environment and Planning B: Planning and Design*, 36 (1), 170–185.

Nichol, J., *et al.*, 2006. Assessment of urban environmental quality in a subtropical city using multispectral satellite images. *Environment and Planning B: Planning and Design*, 33 (1), 39–58.

Nichol, J.E. and Wong, M.S., 2006. 12 Assessing Urban Environmental Quality with Multiple Parameters. *Urban remote sensing*, p. 253.

Noriega, N.F., Soria, C. de la M., 1999. Auditoria Ambiental Municipal: Gestao Local em Cidades Sustentaveis e Saudaveis, *PANGES, Barcelona*.

Norman, J.M., Kustas, W.P. and Humes, K.S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, 77 (3), 263–293.

Nowak, D.J., *et al.*, 1996. Measuring and analyzing urban tree cover. *Landscape and Urban Planning*, 36 (1), 49–57.

Páez, A., Uchida, T. and Miyamoto, K., 2002. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A*, 34 (4), 733–754.

Pao, H.T. and Tsai, C.M., 2011. Multivariate Granger causality between CO2 emissions, energy consumption, FDI (foreign direct investment) and GDP (gross domestic product): evidence from a panel of BRIC (Brazil, Russian Federation, India, and China) countries. *Energy*, 36 (1), 685–693.

Paolini, L., *et al.*, 2006. Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. *International Journal of Remote Sensing*, 27 (4), 685–704.

Pires, S., Fidélis, T. and Ramos, T., 2014. Measuring and comparing local sustainable development through common indicators: Constraints and achievements in practice. *Cities*, 39, 1–9.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. .

Rahman, A., *et al.*, 2011. Urbanization and quality of urban environment using remote sensing and GIS techniques in East Delhi-India. *Journal of Geographic Information System*, 3 (01), 62.

Raphael, D., *et al.*, 1996. Quality of life indicators and health: current status and emerging conceptions. *Social Indicators Research*, 39 (1), 65–88.

Reginster, I. and Goffette-Nagot, F., 2005. Urban environmental quality in two Belgian cities, evaluated on the basis of residential choices and GIS data. *Environment and Planning A*, 37 (6), 1067–1090.

Richter, R., 1990. A fast atmospheric correction algorithm applied to Landsat TM images. *International Journal of Remote Sensing*, 11 (1), 159–166.

Richter, R., 1998. Correction of satellite imagery over mountainous terrain. *Applied optics*, 37 (18), 4004–4015.

Richter, R. and Schläpfer, D., 2005. Atmospheric/topographic correction for satellite imagery. *DLR report DLR-IB*, 565–01.

Rinner, C., 2007. A geographic visualization approach to multi-criteria evaluation of urban quality of life. *International Journal of Geographical Information Science*, 21 (8), 907–919.

Sarmento, R. and Zorzal, FMB and Serafim, AJ and Allmenroedr, LB. Urban environmental quality indicators. *WIT Transactions on Ecology and the Environment*, 39.

Scholars GeoPortal, 2014. Scholars GeoPortal. [http://geo2.scholarsportal.info] [Online; accessed April 19, 2014].

Selçuk, R., *et al.*, 2003. Monitoring land–use changes by GIS and remote sensing techniques: Case Study of Trabzon. *In*: *Proceedings of 2nd FIG Regional Conference, Morocco*, 1–11.

Shoff, C., Chen, V.Y.J. and Yang, T.C., 2014. When homogeneity meets heterogeneity: the geographically weighted regression with spatial lag approach to prenatal care utilization. *Geospatial health*, 8 (2), 557.

Shoff, C., Yang, T.C. and Matthews, S.A., 2012. What has geography got to do with it? Using GWR to explore place-specific associations with prenatal care utilization. *GeoJournal*, 77 (3), 331–341.

Smeets, E. and Weterings, R., 1999. Environmental indicators: Typology and overview. *European Environment Agency Copenhagen*.

Smith, T.E., 1989. Shortest-Path Distances: An Axiomatic Approach. *Geographical analysis*, 21 (1), 1–31.

Snyder, W.C., *et al.*, 1998. Classification-based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing*, 19 (14), 2753–2774.

Sobrino, J.A., Jiménez-Muñoz, J.C. and Paolini, L., 2004. Land surface temperature retrieval from Landsat TM 5. *Remote Sensing of environment*, 90 (4), 434–440.

Sutton, P., *et al.*, 2001. Census from Heaven: an estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, 22 (16), 3061–3076.

Sutton, P., *et al.*, 1997. A comparison of nighttime satellite imagery and population density for the continental United States. *Photogrammetric Engineering and Remote Sensing*, 63 (11), 1303–1313.

Sutton, P.C. and Costanza, R., 2002. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics*, 41 (3), 509–527.

Sutton, P.C., Elvidge, C. and Obremski, T., 2003. Building and evaluating models to estimate ambient population density. *Photogrammetric Engineering & Remote Sensing*, 69 (5), 545–553.

Sutton, P.C., Elvidge, C.D. and Ghosh, T., 2007. Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8 (S07), 5–21.

Szalai, A., 1980. The meaning of comparative research on the quality of life. *The quality of life: Comparative studies. London: Sage.*

Taie, S., ElDeeb, H. and Atiya, D., 2011. A new model for automatic raster-to-vector conversion. *International journal of engineering and technology*, 3 (3), 182–190.

The Mining Association of Canada, 2011. Facts and Figures 2011. [http://www.miningnorth.com/wp-content /uploads/2012/04/MAC-FactsFigures-2011-English-small.pdf] [Online; accessed June 28, 2014].

Topcu, M. and Kubat, A.S., 2009. The analysis of urban features that affect land values in residential areas. *In*: *Proceedings of the 7th International Space Syntax Symposium, KTH, Stockholm*, Vol. 26, p. 26.

Turcotte, M., 2008. Dependence on cars in urban neighbourhoods. *Canadian Social Trends*, 85, 20–30.

United States Geological Survey, 2014. United States Geological Survey. [http://earthexplorer.usgs.gov/] [Online; accessed April 25, 2014].

Van Kamp, I., *et al.*, 2003. Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study. *Landscape and urban planning*, 65 (1), 5–18.

Ward, M.D. and Gleditsch, K.S., 2008. *Spatial regression models.* Vol. 155. Sage.

Weng, Q., 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of environmental management*, 64 (3), 273–284.

Weng, Q., Lu, D. and Schubring, J., 2004. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote sensing of Environment*, 89 (4), 467–483.

Weng, Q. and Quattrochi, D.A., 2006. *Urban remote sensing*. CRC Press.

Yan, W.Y., *et al.*, 2014. Analysis of multi-temporal Landsat satellite images for monitoring land surface temperature of municipal solid waste disposal sites. *Environmental monitoring and assessment*, 186 (12), 8161–8173.

Yan, W.Y. and Shaker, A., 2011. The effects of combining classifiers with the same training statistics using Bayesian decision rules. *International Journal of Remote Sensing*, 32 (13), 3729–3745.

Yan, W.Y., Shaker, A. and El-Ashmawy, N., 2015. Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, 158, 295–310.

Yao, X. and Thill, J.C., 2005. How Far Is Too Far?–A Statistical Approach to Context-contingent Proximity Modeling. *Transactions in GIS*, 9 (2), 157–178.

Yigitcanlar, T. and Lönnqvist, A., 2013. Benchmarking knowledge-based urban development performance: Results from the international comparison of Helsinki. *Cities*, 31, 357–369.

Yue, W., Gao, J. and Yang, X., 2014. Estimation of gross domestic product using multi-sensor remote sensing data: A case study in Zhejiang province, East China. *Remote Sensing*, 6 (8), 7260–7275.

Zavadskas, E., Kaklauskas, A., Šaparauskas, J. and Kalibatas, D., 2007. Vilnius urban sustainability assessment with an emphasis on pollution. *Ecology*, 53, 64–72.

Zha, Y., Gao, J. and Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24 (3), 583–594.