LUMA-GIS Thesis nr 4

APPLICATIONS OF GEOGRAPHIC INFORMATION SYSTEMS AS AN ANALYTICAL AND VISUALIZATION TOOL FOR MASS REAL ESTATE VALUATION:

a case-study of Fontibón District, Bogotá, Colombia

Rafael Villarreal Pacheco

2009

Dept. of Physical Geography and Ecosystems Analysis
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

A Master thesis presented to

Department of Physical Geography and Ecosystem Analysis

Centre for Geographical Information Systems

of

by

RAFAEL D. VILLARREAL PACHECO

in partial fulfilment of the requirements

for the degree of Master in Geographical Information Science

Supervisors:

Dr. Micael Runnström, Lund University

Dr. Armando Blanco Cruz, Universidad Externado de Colombia

ABSTRACT

The majority of approaches for mass appraisal are based on global statistical models that take into account property characteristics, and, in developing countries, location variables are not often explicitly considered. The present study uses Geographic Information Systems (GIS) as an analytical tool to develop a number of location variables and to analyze their impact in the value of properties. Variables like proximity to an airport, proximity to leisure facilities, proximity to shopping centres, and access to main roads, among others, are created, and property values estimated. Surface value maps are created and the effect of the different location variables in the price is hypothesized. Four different statistical approaches are used in order to analyze the effect of proximity attributes: bivariate Pearson coefficient, multiple linear regression, step-wise procedure, and local spatial statistics (Geographically Weighted Regression-GWR) are used to analyze spatially non-stationary behaviours.

The visualization capabilities of GIS prove helpful in constructing value maps, and GIS also proves to be fundamental to report and interpret the results of local spatial statistical models. The results reveal that the spatial variation of some of the parameters is significant and that there are clusters of tendencies. Given this strong clustered distribution, the use of local spatial statistics, to analyze variations of the different statistical parameters across the study area, is ideal. The local statistical results show that prices in expensive clusters seem to be well explained by location factors, unlike prices in less expensive clusters that are poorly explained. The main conclusion from this thesis is that, for the study area, low-priced properties are more difficult to replicate with proximity attributes than high-priced properties.

Keywords: Real estate valuation, mass appraisal, GIS, location variables, local statistics.

SISTEMAS DE INFORMACIÓN GEOGRÁFICA COMO INSTRUMENTO PARA VISUALIZAR Y GENERAR VARIABLES DE LOCALIZACIÓN PARA LA VALORACIÓN MASIVA DE PREDIOS:
caso de estudio Localidad de Fontibón en Bogotá, Colombia

RESUMEN

Los avalúos masivos de propiedades están basados en su mayoría en modelos estadísticos globales que tienen en cuenta las características de las propiedades; sin embargo, los modelos utilizados en Bogotá, no tienen en cuenta variables de localización explícitamente. En este estudio se utilizaron Sistemas de Información Geográfica (SIG) como herramienta analítica para desarrollar una serie de variables de localización y analizar su impacto en el valor de las propiedades; se crearon variables como la proximidad a un aeropuerto, proximidad a centros comerciales y el acceso a vías principales, se generaron mapas de superficie con los valores de las propiedades y se analizó el efecto de las diferentes variables de localización en el precio de las propiedades. Cuatro diferentes enfoques se utilizaron con el fin de analizar el efecto de las variables de localización: coeficiente de Pearson, regresión lineal múltiple, step-wise y se utilizaron modelos locales de estadística (Regresión Geográfica Ponderada o Geographically Weighted Regression GWR) para analizar los comportamientos no estacionarios de las variables.

El SIG como herramienta de visualización demostró ser útil en la construcción de mapas de valor, siendo además fundamental para mostrar e interpretar los resultados de los modelos estadísticos locales. Se encontró que en el área de estudio existe una considerable variación de los parámetros, identificando grupos de tendencias, esto hace que el uso de modelos locales de estadística sea ideal. Los modelos locales mostraron que los valores estadísticos en los grupos de predios más costosos parecen estar bien explicadas por las variables de localización, pero en contraste, para los grupos de predios menos costosos, las variables de localización tienen menos capacidad para explicar la varianza en los valores de los predios.

Acknowledgements

First of all, I am immensely grateful to all the teachers of the program, for the effort they have placed in every course and the high quality of the master, I am incredible happy with the opportunity given to me. I am also thankful to my examiner committee Mr. Per Schubert and Dr. Ben Smith for their corrections and commentaries. Special thanks to my thesis supervisor Dr. Micael Runnström for his opportune commentaries, direction and the time he placed on the thesis. I am also grateful to my supervisor in Colombia Dr. Armando Blanco Cruz who took time from his busy agenda to comment on my work.

En la parte práctica de la tesis utilicé datos Catastrales y sobre criminalidad en Bogotá. Estoy muy agradecido con la directora de Catastro Distrital de Bogotá María Camila Uribe y con Gabriel Vallejo por su ayuda poniendo los datos a mi disposición. Igualmente estoy agradecido con el Observatorio de Convivencia y Seguridad Ciudadana-SUIVD, con el director Rubén Darío Ramírez y Viviana Arango Villegas por sus respuestas oportunas y por facilitarme datos sobre crímenes en el área de estudio.

A number of people also deserve credit for assisting with various aspects of this thesis and the program. Of those that deserve special attention are my friend El Mehdi Oussir for his help in reading, proofreading and commenting parts of the manuscript. Mi papá por su visión, sugerir el tema y su apoyo en teoría catastral y en economía urbana. Mis amigos, Caro y Juanda, que fueron un apoyo importante en el tiempo de estudio en Londres. Finalmente mi mamá, Jime y Ale por estar siempre presentes y pendientes, ellas son mi inspiración.

Rafael D. Villarreal

Contents

List of tables

List of figures

1. INTRODUCTION

Knowing the value of urban properties is one of the main concerns of urban planners, realtors, public authorities, inhabitants and other stakeholders in a city. However, estimating the value of the real estate in a city is far from simple, all of the following should be taken into account during a valuation: national economy fluctuations, market values, construction value of the property, characteristics, and spatial and location factors. An accurate estimate of the value of the lands and properties in urban areas is fundamental for the development of a city because the estimated values are essential for tax purposes. The methodologies for valuating properties must be coherent, consistent, and follow transparent rules. Common approaches for mass appraisal are based on statistical models that take into account property characteristics and few location variables. This study uses GIS as visualization and as an analytical tool to develop a number of location variables, and analyze their impact on the value of the properties by using global statistical approaches (Pearson Coefficient, Multiple Linear Regression and Step-wise procedure) as well as local spatial statistics (Geographically Weighted Regression).

In Bogota, property valuation is mainly done using econometric models (Thompson 2004) based on property characteristics. Homo-value zones are explanatory variables, and property valuations are dependent variables done by on-site valuators. However, for a non-homogeneous and dynamic city of nearly 8 million inhabitants, with 1.833.994 legal properties (Catastro 2005), on-site valuations could be an endless task. Also, the current econometric models used for mass valuation in Bogota do not explicitly consider location variables, and use global approaches that assume the model to be the same for the entire region. In reality, a model of valuation might not be constant over space. For instance, different neighbourhoods might have different model behaviours. Fotheringham et al. (2002) mentions that there are a number of perceptible or imperceptible issues that produce different responses to the same stimuli over space, and different values of the statistics can occur in different locations within the study region.

Aside from the elements mentioned before, the new methods of valuation use computer assisted statistical modelling to perform land and property valuation. These new approaches are called Computer Assisted Mass Appraisals (CAMA), and are frequently integrated into GIS for location analysis and for visualization purposes. As Rodriguez et al.(1995) state, the importance of location has long since been recognized in property valuation, but before the development of GIS, it was difficult to accurately measure the impact of location in real estate models.

This thesis has various purposes. One is to explore and review literature in mass valuation methodologies and its relationship with GIS. A second is, in the context of Bogota, to use the spatial capabilities of GIS to study the impact of location and spatial variables in the values of residential properties. Finally, this thesis aims to apply local statistical modelling with GIS to mass real estate valuation and compare the results with the traditional global approach.

1.1 Objectives

This thesis has theoretical and practical objectives. The theoretical objective is to do a literature review of valuation methodologies and cases of study, and from there, explore the possibilities of the use of GIS as a complement to the econometric models of land and property valuations, specifically in the following practical ways:

1- Use of the spatial capabilities of a GIS to create location variables that may have an impact in the value of residential properties.
2- Use of GIS as a visualization tool to create value maps and analyze and interpret the value patterns and the possible effect of each of the location variables in the value of the residential properties.
3- Use the location variables created, to estimate global statistical models of property and land value. Analyze the correlation, the weight and the statistical significance of different location variables.
4- Estimate a local spatial statistical model and explore the results using the visualization capabilities of GIS.
5- Compare the three approaches: the visual interpretation, the global statistical analysis and the local spatial statistical analysis.

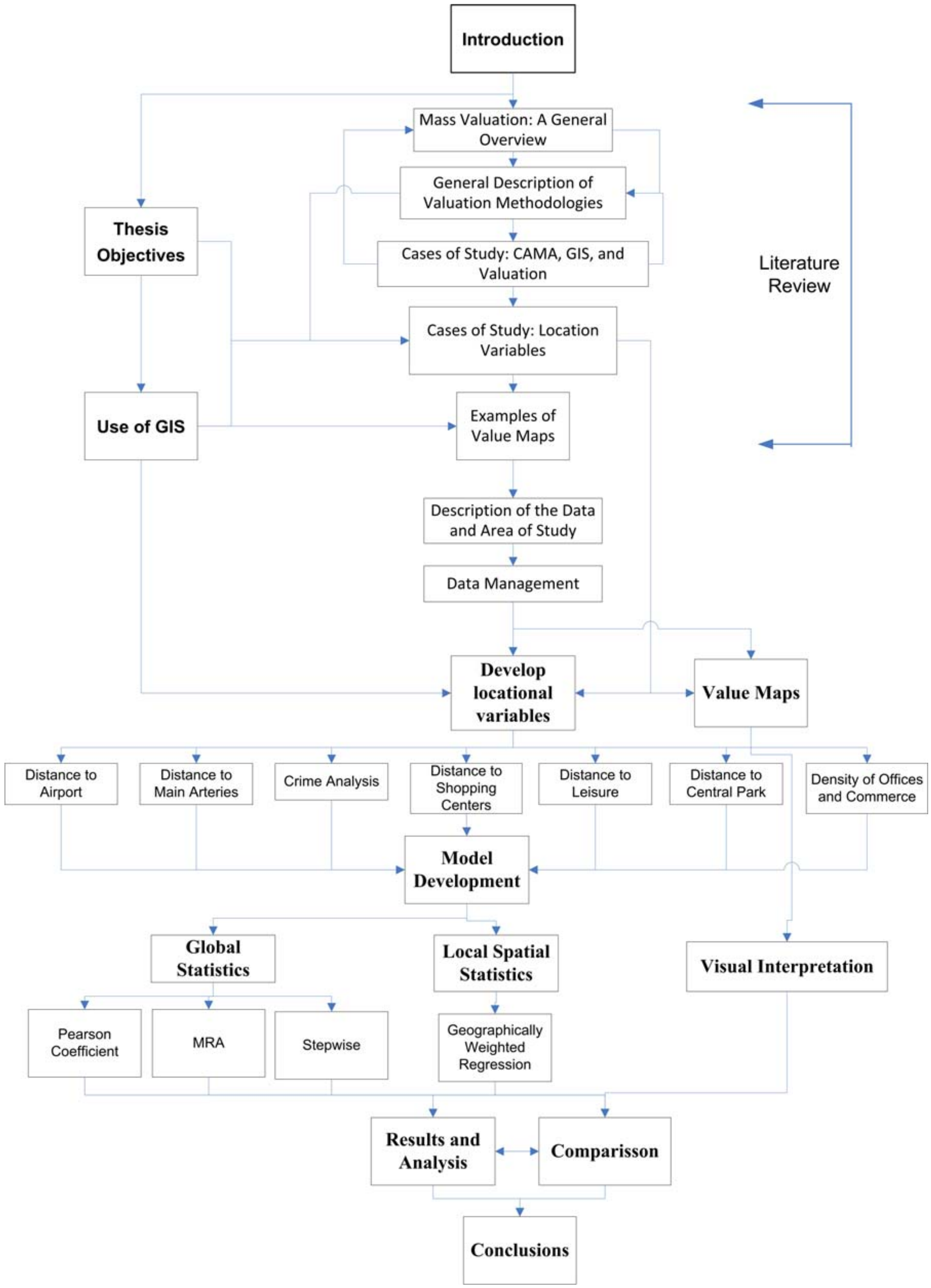

Figure 1-1 Flow chart with the structure of the thesis

1.2 Background and Context

The valuation process in Colombia is done through the regional or national offices of Cadastre. The national agency, Geographic Institute Agustin Codazzi (IGAC), has a direction of Cadastre, which is accountable for the administration of the Cadastre for most of the nation, except in the cities of Bogota, Cali, Medellin, and the state of Antioquia (Bustamante and Erba 2008).

The focus is on the city of Bogota; therefore, most of the data used will come from the Cadastre of Bogota. The Cadastre of Bogota was created in 1981 and became fully operational in 1991. It is an autonomous administrative department, only dependent on the office of the Major. At first, the cadastral information was not updated, until a major initiative from the Mayor of the city allowed for the update of the 1,734,622 properties in 2000 – 2003, which was the totality of the legal properties the city had at that time. This means that in the year 2003, the information that the Cadastre of Bogota had in its digital map was similar to that of a real city. This update directly impacted the income of the city's due property tax, because it was based on a more accurate number of properties valuated, which was estimated to be U.S.$24 million per year (Bustamante and Gaviria 2004). The information about properties is updated every five years, taking into account information regarding physical changes in properties, productivity changes, public works, and local real estate market variations (Bustamante and Gaviria 2004).

Having updated information about the characteristics of each property is important for the finances of the city. From this information it is possible to estimate the value for each property in the city. This process of valuation can be performed in many different ways, as we will be able to see in the review later. In Bogota particularly, this process is a mixture of econometric models fed by on site information from professional appraisers, and a self valuation variable. About this Thompson (2004) mentions:

> *The econometric model took into consideration typical assessment variables but also considered a key element in the Bogotá cadastre, the 'public value estimate.' … The property owner or occupant provides an estimate of property value and its depreciation or appreciation as required by the Unified Property Tax Reform Act.*

The objective of this self valuation is to simplify the process of valuation until a team of professional appraisers are able to valuate the property.

Regarding the use of GIS, Thompson (2004) also mentions that its use is fundamental for the Department of Cadastre's process of updating. The databases are growing in size, and in the future there are plans to offer interactive access to some tax information.

Despite the recent advances, there is still a long way to go. The valuation methods still need to be improved, especially the mass valuation methods. On site valuation is prohibitively expensive, so more cost effective techniques must be obtained to update the valuation of properties. The Cadastre needs to reduce the on site work by increasing modelling capabilities. However, this is far from simple, especially in a developing non-homogeneous megalopolis, where formal information about real estate is available for only part of the transactions taking place.

1.2.1 General Information about Bogota

Bogota is the capital, and most important city, of Colombia. It has the largest economy in Colombia and houses all the powers of the Republic: Congress, Supreme Court, President and Military Headquarters.

The main authority in the City is the Mayor, who is elected every four years.

Despite being a city in a developing world, Bogota is a city with a wealthy economy and important governmental investments. Like most of the developing cities, Bogota is a city of contrasts, which includes both wealthy and extremely poor areas. The main economic activities are services, such as banking, insurance sales, stock market trading, engineering, and consulting, and it also has the largest industrial facilities in the country. In addition, Bogotá houses the main educational institutions in the country.

The 2005 census shows that the city has a population close to seven million with just over eight million counting the surroundings (DANE 2008) and 1.833.994 legal properties (Catastro 2005). The city size is particularly interesting, with 1,587 square kilometres of urban area, making it one of the largest cities in the world; it is similar in size to Greater London and Mexico City.

The city is located in the center of Colombia, which is a strategic geographic location. Bogota is politically organized into districts called localities; there are 20 of them as follows: 1. Usaquen, 2. Chapiero, 3. Santa Fe, 4. San Cristobal, 5. Usme, 6. Tunjuelito, 7. Bosa, 8. Kennedy, 9. Fontibon, 10. Engativa, 11. Suba, 12. Barrios Unidos, 13. Teusaquillo, 14. Los Martires, 15. Antonio Nariño, 16. Puente Aranda, 17. La Candelaria, 18. Rafael Uribe Urie, 19. Ciudad Bolivar, 20. Sumapaz.

Each of the 20 districts has a Mayor designated by the Principal Mayor of the city and some of the localities are of the size of a city with over a million inhabitants.

2. VALUATION METHODS

This chapter presents a general overview of the valuation process, defines mass valuation, and some of the most important methodologies for mass valuation. It also discusses the advantages and shortcomings of the main approaches.

2.1 Mass Valuation a General Overview

Land and property valuation systems are generally administered by national or regional authorities; they are strongly linked as part of the fiscal system, and use land cadastre data for taxation and other purposes.

In the case of Bogota, the task of valuation and cadastral administration is done by the Department of Cadastre, and the fiscal system is the responsibility of the Secretaria de Hacienda (Secretary of the Treasury). These two entities must have a close relationship, especially because of their joint responsibility for property tax, estimation, and collection.

The need for mass valuation became an important issue when governments and authorities started to apply property tax that required the valuation of a large number of properties at the same time. The on site valuation process is extremely time consuming and expensive, especially for large urban areas. A mass valuation appraisal is supposed to be homogeneous, transparent, accurate, fair, and relatively inexpensive, while offering information for every property (Federal Land Cadastre Service of Russia 2001).

A number of authors and reports, for instance Eckert (1990), Malme (2004) and The Federal Land Cadastre Service of Russia (2001), define mass valuation as a systematic valuation of groups of real estate units, performed on a certain date with the help of standard procedures and statistical analysis. It is usually performed for taxation purposes.

In contrast, individual valuation is an estimate of the value of a property of known characteristics as of a specified date. It is usually an on site process performed by a professional appraiser, but there are new methods that rely on statistical models and on site information. As can be seen later in this thesis, the frontier between models of valuation and models of mass valuation is unclear, especially with later advances in statistical modelling and computer technologies. As McCluskey et al. (1997) state, mass appraisal differs from single property appraisal only in terms of scale. Today, with the availability of data, expert support, and strong computer equipment, it is possible to build more complete mass appraisal models that are close in results to individual models.

There are different approaches for land and property valuation, yet they are generally divided by traditional and advanced approaches. These approaches are reviewed in the next section, but in general the most effective mass valuation models are obtained using a combination of approaches.

2.2 Valuation Methodologies

Valuation methodologies have evolved greatly in the last 30 years. In the past, they were mainly based on manual valuations performed by professional appraisers, and the concept of valuation models was nearly nonexistent. Today, even the manual valuation relies on complex databases, market values, statistical models, and GIS.

The development of valuation models in recent times is due to a variety of factors. The first reason for this development is institutional, which is the necessity for comprehensive valuation models for taxation purposes as explained previously. The second reason is the availability of historic market data that can now be stored in computers. The third reason is the advances in statistical theory and techniques, and the fourth is the advances in computer processing that have allowed for the development of powerful data management software, statistical packages, and GIS. The combination of these advances has helped the development of the processes of appraisal of real estate. The new methods to perform real estate appraisals are getting less dependent on subjective judgement and are increasingly adopting the scientific foundations based in statistical and econometric models (Vazquez 2006).

There are a variety of methods for real estate or property valuation. The most common are the following three traditional approaches: Market Value Comparison or Comparable Sales, Cost Analysis, and Cost of Developments. Generally a combination of the three with the use of statistical regressions is used.

Real estate consists of various elements, Fisher and Martin (2004) divide the physical elements into two major components: the land and the improvements that are added to the land. There are also legal property rights associated to the physical part, which are defined in Fisher and Martin’s book as the interest benefits and rights inherent in the ownership of a property.

Fisher and Martin (2004) mention a number of real estate studies that are part of the valuation process; a few of them are: Market Study, Marketability Study, Investment Analysis, Feasibility Study, Environmental Impact Study, Cost-Benefit Study, and Land Utilization Study. These studies are generally done for individual property valuation, but with the advances of mass valuation, complex models might be able to include some of these studies in related variables.

Various authors (German et al. 2000; Pagourtzi et al. 2003; Vazquez 2006) divide the valuation methods into two major groups; They are the traditional methods and the new or advanced approaches. Both groups share the principle of comparing market value; the difference is that in the first group the comparison is done through direct comparison or relies on the experience of an appraiser, who determines the value of the property, and is less statistically strict. Pagourtzi et al. (2003), defines the second group as methods that simulate the market and the decision making process in order to estimate the prices; these methods are more mathematically demanding since they are more statistically strict and use advanced statistical models.

Pace (2001) states that a dramatic change in valuation technology is the application of CAMA (Computer Assisted Mass Appraisal). The use of this technology has improved the accuracy of mass appraisals and reduced their costs. But, what is inside of the black box? The theory about the different methods is reviewed in this chapter.

The adoption of either of the methods and the complexity of the statistical model depend on several factors, including the availability of data, the availability of technical infrastructure for data processing (software and hardware), the availability of specialized personnel (e.g. expert statisticians), and the popularity and widespread acceptance of either procedure, among other factors.

2.2.1 Traditional Methods of Valuation

Different authors have different names for the approaches in this group, but at the end they are the same but have superficial differences.
Fisher and Martin (2004) mention three approaches: the Cost Approach, the Direct Sales Comparison approach and the Income Capitalization approach.
For instance German et al. (2000) mention as traditional methods, Comparable Sales, Income Analysis, Cost Analysis, and Cost of Developments.
Pagourtzi et al. (2003) have a similar list of methodologies: Comparable method, Invest/Income Method, Profit Method, Development/Residual Method, Contractor's Method/Cots Method, Multiple Regression, and Stepwise Regression Method.

The most representative of traditional methods will be briefly discussed, but it is worthy to point out that they are usually used in combination.

2.2.1.1 Comparable Sales Method or Comparison Approach

The Comparable Sales Method also known as the Market Approach, is probably the most common method of valuation, since it takes into account the market value of similar properties with similar characteristics. This method is used for a multitude of purposes of valuing real estate, including the cadastral activity.

The value of an individual property is estimated by direct comparison to similar properties that have recently sold. The idea behind this method is that the value of a property is closely linked to the sale prices of similar properties in the same market segment (Fisher and Martin 2004). However, since there are not two properties with the exact same characteristics, the main task is to determine the differences between similar properties and how these differences affect the price.

For mass valuation some level of detail has to be sacrificed, since it is not possible to account for all the attributes of a property, but in general this method can be used in combination with regression and other methods.

Pagourtzi et al. (2003) divides the method in three main stages:

1. Find the sales data of buildings that are highly comparable with the property to be valued.
2. Adjust prices of the real estate taken as reference to the characteristics of the objective building.
3. Estimate the market value by using different techniques.

This type of model relies on complete, timely, and accurate market data, as well as data series of transactions over time. This data can come from different sources like the cadastre office, real estate agencies, and government offices where the transactions are stored, among other sources.

Fisher and Martin (2004) mention that in cases where there is poor information about recent sales, or unorganized market values, this method might not be reliable and its application must be cautious.

2.2.1.2 Cost Analysis Method

The cost analysis method has nearly no connection with market information. This valuation approach consists mainly of determining the value of the property from the sum of its components. Fisher and Martin (2004) state that with this method the value of a property is estimated by adding the contributing value of any improvements in a land to the value of the vacant land.

This method assumes that properties can be worth no more than their cost of construction and the remaining value is obtained from the improvement to the parcel and to the land (German et al. 2000).

According to Pagourtzi et al. (2003), this method is effective in cases where the property has very detailed specifications that make it impossible to rely on actual market data to know its value.

This method, sometimes called the add-sum method, involves a high correlation between construction cost and price, and is precise for the estimation of the real construction value. Its applicability for mass valuation is very limited because it requires a lot of detailed information about each property. Nevertheless, it can be useful for mass appraisal when it is used in conjunction with market values, and it offers a more realistic view of the value.

2.2.1.3 Income Analysis Method

This method is also known as the Income Capitalization Approach. The value of the property is derived by converting the expected income generated from a property into a present value by using income capitalization procedures (Fisher and Martin 2004).

In this approach the property is viewed from an investor's point of view. It begins with an estimate of the income yielded by the development of the property, in the form of the resale, rent, or expected future cash flow. Then, the value of the property is calculated. This is a direct comparison approach, in which the estimated value assumes that investors will earn a rate of return consistent with the availability for alternative investment at comparable risk (Fisher and Martin 2004).

This is not a common method to be used in large scale valuations, but it is important because it is a common complement for some of the other valuation approaches.

2.2.1.4 Multiple Regression Analysis Method

Strictly speaking, this is not a valuation method, but it is more a tool for valuation that because it is so commonly used it has became a method. This is probably the reason that Pagourtzi et al. (2003) has included this method as a traditional approach.

Multiple regression analysis is the traditional linear regression model, where there are dependent and independent variables. The dependent variables are explained when the independent variables are multiplied by fixed unknown parameters and an offset is added, which have to be estimated, and are also explained by a normal distribution error term with a mean zero and a variance of σ^2.

Linear regression has a number of strong assumptions:

- The explanatory variables are uncorrelated.
- The error terms have to be uncorrelated and the variance of the error terms is constant.
- All the error terms have a mean zero and a variance of σ^2, and have a normal distribution.

The objective of the multiple regression model is to estimate the implicit market price of each property by using linear regressions of transaction prices or valuation prices, compared with attributes and measurements that reflect the presence of certain features of the property.

In general, this type of model represents a dependent variable V (value), which varies according to independent or explanatory variables that characterize the property. These attributes can be, for instance, size, type of construction, number of bedrooms, and other location attributes like neighbourhood type or accessibility.

Let's say that the attributes are A_1, A_2 ,..., A_n. The multiple regression equation is of the form**:**

$$V = f\left(A_1, A_2, .., A_n\right) \tag{2-1}$$

With an equation of the form:

$$V_i = \beta_0 + \sum_{i=1}^{n} \beta_i A_i + \varepsilon \tag{2-2}$$

V is the price or value of the property, β_o is a constant term, the A's are the attributes, the β's are the estimated weights associated to each attribute, and the ε is the error term or residuals.

The theoretical support of this methodology is amply discussed in statistical books. For real estate valuations and for urban economy, according to Lentz and Wang (1998), the most common used technique is the Ordinary Least Square Method (OLS), which minimizes the sum of the squares of the errors between the predicted values of V and the known values of V. However, most of the cases that use this approach do not fully obey the assumptions of linear regressions. For instance, it is well known and studied that the distribution of the error term is typically heteroskedastic, meaning that the residual term has different variances, failing the OLS assumption of constant variance in the error term. For spatial variables the errors are typically spatially correlated, failing the assumption of independent and identically distributed variables. A number of papers deal with this problem of strong regression assumptions, some of them, related to valuation regressions, will be briefly reviewed later on the advanced methods section.

The multiple regression method has been demonstrated as being the primary technique used in the mass appraisal of real estate (McCluskey et al. 1997). Most of CAMA models rely on Multiple Regression Analysis (MRA), which is indicated in the table below from McCluskey et al. (1997).

Table 2-1 CAMA applications in the World

Country	Department	Model
Australia (Queensland)	Department of Lands	MRA
Sweden	National Tax Board	MRA
Northern Ireland	Valuation and Lands Agency	MRA, ANN, CSA
Tasmania	Valuer General's Office	MRA, AEP
New Zealand	Valuation New Zealand	MRA
Singapore	Singapore Valuation Authority	INDEXATION
Hong Kong	Rating and Valuation Department	MRA, INDEXATION
Malaysia	Valuation Division	EXPERT
USA	Assessment Offices	MRA, AEP, CSA
British Columbia	BC Assessment Authority	MRA

Notes:
MRA = Multiple regression analysis; ANN = Artificial neural networks; CSA = Comparable sales analysis; AEP = Adaptive estimation procedure; EXPERT = Expert system; INDEXATION = Indexing of existing assessments

2.2.2 Advanced Methods

Advanced methods are methods developed to overcome the shortcomings of traditional methods, lack of data, theoretical assumptions, global models, and problems with the assumptions of linear regression. The methods selected for this section are divided in theoretically advanced and computational advanced; the latter is the use and integration of CAMA with GIS.

Pagourtzi et al (2003) list the advanced methods as follows: artificial neural networks, hedonic pricing models, spatial analysis methods, fuzzy logic, and autoregressive integrated moving average. For advanced methods, an ample choice of statistical methods can be considered; for instance, Vazquez (2006) adds to this list an interesting method called Rough Set Theory.

This part of the review is organized as follows: first, a review of the Rough Set Theory approach; then, the Spatial Analysis Methods and Local Statistics (Geographically Weighted Regression), and the Econometric Methods (statistical approaches related to multicollinearity, heteroskedasticity and spatial econometrics), and finally, a brief introduction to CAMA.

2.2.2.1 Rough Set Theory Method

Rough set theory is a rule based approach created to analyze imprecise information full of attributes. According to Pawlak and Slowinski (1994), who are the principal developers of this method, the imprecision in the information may be caused by the granularity of representation of the information. Rough set theory identifies the objects that have similar properties, are indiscernible in terms of descriptors, and consequently, are treated as identical or similar.

In the context of property valuation d 'Amato (2002) and d'Amato (2004) have proposed the use of rough set theory as an alternative to overcome the limitation of traditional regression models, especially in markets with high uncertainty due to the unreliability of information sources. He states that in property markets where the dynamic of the price could be described through non-monotonic processes, rough set approach can be used.

In conclusion, d'Amato (2002) states that Rough Set Analysis (RSA) could represent another useful appraisal method in those markets where the mathematical system could be ineffective. *"The RSA allows the appraiser to define deterministic rules between the property features and their price. The property appraisal is a consequence of these rules,"* and finishes by stating that probably the best application of the RSA is the mass appraisal.

Rough set theory can be an interesting method of valuation, which needs more study, since there are just two cases of study in the field of property valuation, and both of them are from the same author. These studies are not enough to prove that it is really a feasible option that is superior, in specific cases, to traditional regression models.

2.2.2.2 Spatial Analysis Methods

Pagourtzi et al. (2003) mentions the use of GIS to improve the measurement of location and accessibility variables, but also that its analytical functions can be used to perform more advanced spatial analysis, involving advanced spatial statistics.

Anselin (1992) analyzes the possibilities for spatial statistics with GIS and Fischer and Nijkamp (1992) mention some limitations of GIS for modelling that restrict the spatial analysis possibilities; limitations that today have been overcome with the integration of GIS with other platforms.

Spatial analysis methods include the development of spatial econometrics to perform spatial autocorrelation analysis, according to Dubin (1988), Anselin (1990), and Dubin (1992); or it includes the use of GIS with statistical techniques to determine optimal distances to determined facilities, according to (Des Rosiers et al. 1996; Des Rosiers et al. 2000). Other spatial analyses, such as spatial interaction models, are often referred to as gravity models to forecast traffic flow, store patronage, and shopping center revenue, and may be used to identify optimal site location (Rodriguez et al. 1995).

Spatial analysis and GIS allow for the creation of a number of advanced spatial variables, and for the analysis and econometric improvements to traditional models. Pagourtzi (2003) states that the research in the area of GIS and advanced spatial analysis has contributed to a better understanding of the measurement of location effects.

2.2.2.3 Local Statistics Modelling

The local forms of spatial analysis are a very rare exception to the overwhelming quantity of global analysis that dominates the literature (Fotheringham et al. 2002). These methods range from running traditional regression for different spatial units, in this case they can be sub-districts or neighbourhoods, to a sophisticated method such as the Geographically Weighted regression (Fotheringham et al. 2002).

The objective of these approaches is to analyze regional variation in spatial relationships and how processes vary over space. These types of spatial regression models estimate local statistics parameters, which differ from the single global statistics of the traditional models. Local statistics try to solve the problem of the relationship between variables and its location in space. In the spatial regressions, the data points are weighted according to their distances between each other. Therefore, the relationship between dependent and independent variables vary depending on the location of each data point; in simple terms, a model is estimated for each point.

Fotheringham et al. (2002) mentions a number of important differences between traditional global models and the local statistics approach. In terms of outcomes, global statistics produce single values, for instance one mean value, a standard deviation, and one coefficient per variable. On the other hand, local statistics have multiple values, since its spirit is the idea that different locations can have different values of the statistics. Consequently, the value of each local statistic is a measure of the attribute or the relationship being examined in the surroundings of a location within the study region. This yields to a second big difference, which is that local statistics need to be mapped to take full advantage of the approach, since each local parameter or statistical value can be shown in a map to analyze the variability over space and patterns, and GIS is an ideal tool to map local statistics results,.

Summing up, Fotheringham et al.(2002) declares:

> *Local statistics emphasise differences across space, whereas global statistics emphasise similarities across space. Global statistics led to the idea that all parts of the study region can be accurately represented by a single value, whereas local statistics can show the falsity of the assumption by depicting what is actually happening in different parts of the region.(7)*

A summary of the main differences between global and local statistics is presented in the table below:

Table 2-2 Differences between local and global statistics

Global (e.g. Traditional Multiple Regression)	**Local** (e.g. Geographically Weighted Regression (GRW))
Summarize data for the whole region	Local desegregations of global statistics
Single-valued statistic	Multi-valued statistic
Non-mappable GIS-unfriendly	Mappable (GIS-friendly)
Aspatial or spatially limited	Spatial
Similarities across space (regularities)	Emphasise differences across space

Modified from: Fotheringham et al. (2002, p.6)

An important concept to consider when studying local statistics is the spatial non-stationarity. Fotheringham et al.(2002) describes it as a process which is not constant over space, and any relationship that is not stationary over space will not be represented particularly well by a global statistic. Property and land values vary across space; the value of a house is clearly spatially non-stationary.

Geographically Weighted Regression (GWR)

One of the most renowned methods for local statistics is Geographically Weighted Regression (GWR), developed by (Fotheringham 1997; Fotheringham et al. 1998; 2002). Its success relies on the fact that it is based on the traditional regression approach, but incorporating local spatial relationships into the regression structure.

In GWR, a regression is estimated for each point, and each data point is weighted by its distance from the regression point, thus, points closer to the regression point have more weight (the maximum weight is when it shares the same location) in the local regression than the points that are farther (Fotheringham et al. 1998; 2002). For that reason each point will have a unique estimation.

An important part of the GWR is that the weight decreases continuously as the distance between the two points increases, for that reason the concept of bandwidth is important, a measure that has to be chosen in the regression. Bandwidth is a measure of the distance-decay in the weighting function and indicates the extent to which the resulting local calibration results are smoothed.

In Fotheringham et al. (1998; 2002) the traditional equation for linear regression is extended by allowing local rather than global parameters to be estimated, rewriting the linear regression equation in the following way:

$$V_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) A_{ik} + \varepsilon_i \tag{2-3}$$

(u_i, v_i) are the coordinates of each point in space. This equation recognizes that spatial variations in relationships might exist and provides a way in which they can be measured.

The ideal way to see the results of GWR models is through surface maps of parameters estimated, as shown in the results in Chapter 6 of this thesis.

It is important to point out that advanced possibilities for GWR, like fixed spatial kernel or adaptive spatial kernel, and the specific estimation procedures, are not explained because they are out of the scope of the thesis.

For residential property valuation, local statistics has been used in a number of cases. For example, the results of Pavlov's (2000) study of house prices in Beverly Hills, are presented as surface maps and 3D surfaces. Fotheringham et al. (2002) study the house prices in London by using the price data of 12.493 properties and by comparing a global regression model and a GWR model. The estimates are presented as surface maps, and per borough. Hernandez et al. (2003) develop a residential price model for the Greater Toronto Area by applying GWR to address the local variations that are hidden within the traditional global statistics; they present the results as surface maps, but state that surface generation requires careful consideration of the most appropriate methods for dealing with interpolation issues.

2.2.2.4 Econometric Methods

There are several cases of study where advanced econometric models have been used for real estate valuation. The main objective of most of the econometric models is to overcome the strong assumptions of the ordinary linear regression model.

The assumptions are violated especially when the properties are homogeneous, in these cases the distribution of the residual term is usually heteroskedastic and the errors are usually correlated, failing the assumption that hold linear regression theory. It is important to mention that these problems exist, and that there are some ways to test, make more flexible, and/or force the models to comply with the assumptions. This problem is explained by Basu and Thibodeau (1998)

> *Hedonic house price parameters are usually estimated using procedures that assume independent observations (such as ordinary least squares). If hedonic residuals are spatially auto correlated, the resulting parameter estimates will be inefficient and will produce incorrect confidence intervals for estimated parameters and for predicted values.*

One of the simplest econometric approaches as explained in Pagourtzi et al. (2003), is a modification to the multiple regression model that is called stepwise regression model. The objective of this method is to minimize the multicolinearity. The multicolinearity is almost inescapable and probably one of the main violations to the linear regression assumptions. To overcome the effect of multicollinearity, the stepwise regression method consists to test the impact of the introduction or exclusion of each of the variables; this method is also called the five-step method. This approach has be used for real estate valuation according to Des Rosiers et al (2000).

In Des Rosiers et al (2000), the specific property attributes are first introduced and then location attributes, like proximity and neighbourhood characteristics, are added. In conclusion, despite the fact that the performance of the model is high and the collinearity is controlled, there will also be remaining autocorrelation in the error terms.

Dubin (1988) states that spatial autocorrelation is one of the main problems of spatial statistics. It happens when the properties are close and share location characteristics. There are usually two approaches: ignore it or assume it. He provides a third alternative that consists of estimating the covariance matrix of the error terms so that this information can be used to obtain an efficient estimate of the regression coefficients and unbiased estimates of the standard errors.

In conclusion, Dubin finds that by ignoring spatial autocorrelation, the OLS estimates are inefficient because the OLS estimates of the standard errors are biased and the predicted values may be inaccurate. The model developed by him, which includes neighbourhood or spatial variables shows that a proper choice of independent variables can reduce the amount of spatial dependence in the data. The use of spatial variables might help in reducing the autocorrelation but will not necessarily fix the problem. This is because spatial variables are generally measured over arbitrary spatial boundaries.

In a second study, Dubin (1992) states that most of the hedonic estimations show few significant coefficients on the neighbourhood and accessibility variables. The approach he took consists in omitting all neighbourhood and accessibility measures from the set of explanatory variables and instead estimates the resulting autocorrelation in the error term. With this technique he found that there is a portion of the price due to structure that can be calculated by multiplying the vector of attributes by the estimated coefficient, but there is penalty or premium due location, given in terms of neighbourhood and accessibility.

Basu and Thibodeau (1998), in the context of Dallas, analyzed the spatial autocorrelation in house prices by taking into account prices of one type of house. They mentioned that house prices are spatially correlated because of two reasons. First, neighbourhoods tend to be developed at the same time, and tend to have similar characteristics, and second, neighbourhood residential properties share location amenities, for instance access to shopping centres, schools, etc.

Since hedonic property valuation regressions seek to estimate variation in house prices using property structural and location characteristics, the error terms produced by these regressions are usually spatially correlated (Basu and Thibodeau 1998). They compared OLS and kriged estimated generalized least squares (EGLS), finding the latter more accurate in spatially auto correlated submarkets.

In their study, Fik et al. (2003), vary the coefficients over space because of the interactions between the absolute location variables and structural attributes. The results were acceptable in terms that a number of spatial coefficients were significant, indicating the presence of spatial heterogeneity in the prices of these structural attributes. However, there is a limitation because they only used three structural attributes. In conclusion, they suggest that the value of location may not be separable from other determinants of value.

Tse (2002) develops a stochastic approach, making the unobservable variation to follow a probabilistic distribution. With this procedure, the author was able to correct the autocorrelation bias generated in the traditional OLS models.

A common approach to deal with spatial heterogeneity is to define geographic areas with similar characteristics as submarkets, and estimate models for each submarket region. Unfortunately, these housing submarkets are not easy to delineate and require extensive work in the case of mass valuation models. The other deficiency is that they force the price to be a discrete variable over space.

2.2.2.5 CAMA-GIS Interaction

As mentioned previously, a number of authors, such as (German et al. 2000; Pace et al. 2001), consider that one of the greatest changes in the assessment practice is the use of CAMA. McCluskey et al. (1997) suggests that the interaction between CAMA and GIS introduced new considerations to take into account as a guide in the process of valuation.

CAMA, in definition, is the use of hardware, software, and mathematical models to establish relationships between property characteristics and its value in order to achieve a standardized calculation of property values. CAMA models are closely linked to the use of GIS. CAMA integrated with GIS offers innumerable possibilities for land valuation and other cadastral purposes; not only the integrated visual capabilities of GIS, but also possibilities to generate spatial variables.

CAMA, when used for mass valuation purposes, uses any of the previous approaches mentioned, but the most common is the multiple regression method with some econometric amendments.

2.3 Chapter Conclusions

In this chapter the general methods for valuation were reviewed, starting with the traditional methods (the cost approach, the direct sales comparison approach, and the income capitalization approach) and then the advanced methods, in particular the econometric approaches. Cadastral authorities usually follow the fundamentals of any of the three traditional approaches with regression analysis.

The adoption of either of the methods and the complexity of the statistical model depend on several factors, including the availability of data, the availability of technical infrastructure for data processing (software and hardware), and the availability of specialized personnel (e.g. expert statisticians).

The most common approach for mass valuation is the MRA, and there a several statistical approaches to overcome the limitation imposed by its strong assumptions. The assumptions are violated especially when the properties are homogeneous, in these cases the distribution of the residual term is usually heteroskedastic and the errors are usually correlated. For spatial variables, the errors are typically spatially correlated. Some of the approaches that account for these problems are presented in the advanced econometric approaches.

An interesting approach is Local Statistic Modelling. These approaches analyze regional variation in spatial relationships and how processes vary over space. This type of analysis seems to be worthy in real estate, since property values and the impact of each variable appears to be spatially non-stationary, and traditional global statistics models ignore potential variations over space.

The state of the art instrument for valuation is the integration of CAMA and GIS, which will be reviewed in detail in the next chapter.

3. REVIEW OF CASES OF STUDY, GIS AND LOCATION VARIABLES

This chapter presents cases of study. The first section is about GIS and CAMA in valuation; this is important because it gives an idea of the type of models and variables used in the real world. Then, in the second part, a review on location variables is presented, which is relevant because it gives an overview of the most common location variables and their importance in valuation models. This will be used as the foundation to develop the location variables of this thesis. The third part of this chapter is GIS as visualization tool; it is a short section reviewing how GIS has been used for visualization in real estate valuation and its importance.

3.1 CAMA, GIS and Valuation

There has been a long relationship between GIS and real estate. GIS has mostly been used as a visualization tool to examine and store the characteristics of properties. In recent years, with advances in GIS capabilities, it has been largely used for property market analysis and real estate property management. For an ample description of cases of study of applications of GIS in property management see the book by Wyatt and Ralphs (2002).

Today, along with other statistical packages, GIS is an important part of the valuation process around the world. GIS can be tailored to specific applications, like with CAMA models, where the databases are strongly linked to GIS in order to provide location variables and as a visualization tool. In general terms, GIS is ideal for examining the spatial component of real estate since it can be used to produce location variables that can be applied to a variety of real estate analyses.

This section is an extension of the CAMA and GIS integration presented above. It is a review of cases of study of valuation models, where GIS and CAMA have been used.

McCluskey et al. (1997), describes the application of a CAMA model within a GIS, to valuate residential properties in Northern Ireland. They make emphasis in the importance of data, and review a number of mass appraisal methods. They conclude that the application and use of a particular modelling technique depends largely on data availability and data characteristics. In their model they apply Multiple Regression Analysis in association with a computer based comparable selection technique, known as comparable sales analysis (CSA).

The application of GIS to this mass appraisal model is for two reasons, as complement to the model by reflecting the location of the wards, and most importantly, it is applied to enhance the quality assurance process in terms of visually analyzing the pattern of predicted values. The authors conclude with the importance of GIS to display property values, which has notable advantages over standalone mass appraisal systems.

Milam and Mitchell (2003), explain how a fully functional CAMA-GIS model supports the field appraisers. They use GIS applications to get some extra information for the appraisal process. "*Powerful GIS functions that allow visual analysis of location factors that influence property values are only accessible in the office and not in the field where they would best be used while visiting the property.*" This is an example of the use of CAMA-GIS integration to support the individual appraisal of properties. The appraiser is able to do part of the work from the office, by checking orthophotos, analyzing the specific characteristics of each property, and spatially viewing the area in order to analyze the characteristics of the surrounding location. This case is an example of how mass appraisal methods can complement individual valuations.

Welcome and Clark (2003) developed an Automated Valuation Model (AVM), that offers a detailed property valuation. It uses the fundamentals of mass appraisal models to perform an automated valuation with a good level of detail. They develop a single model for the entire jurisdiction, which is a linear regression model with 9 variables that is calibrated daily using prices of sales. The variables represent size, quality of construction, age of improvements, market changes, and as a location variable a variable with the linear ranking of each neighbourhood. The authors mention that in recent years, the changes in value are more due to market changes than to properties attributes.

As a result the model produces:

- A point value estimate that is displayed as part of the AVM output.
- Coefficients for each of the model variables, which are stored in their own file and used to adjust the selling prices of the comparables.

The final result of the AVM is a report for the user, indicating the valuation process carried out and the valuation results that include:

1. a map identifying only the subject parcel
2. a map identifying the subject in all five comparable sales, and
3. a map shaded to reflect the presence of flood plains near the subject

This AVM requires constant inspection of properties and offers specific information, which makes it suitable for small towns, but for mass valuation in large cities is not practical. In this approach GIS is used to establish neighbourhood boundaries, distances, and proximity to other properties in order to compare sale prices. It is integrated to the CAMA to be used mainly as a visualization tool.

Gloudemans and Montgomery (2006) is an interesting case of study in Northern Ireland. They developed a valuation model, using a statistical package for analysis and then imported them into GIS in order to visualize the results. In their integration of CAMA and GIS, as input for the model, they used market price of properties from 2001 to 2004, divided the study area in 25 market areas, classified the properties into Urban Suburban, Rural Village, or Rural, and they had a clear idea of the location of primary roads and secondary roads. They exported data from GIS to be modeled in statistical software, and then imported the results back into GIS to be visualized. The statistical package was used to perform data management, exploratory analysis, and statistical modelling. Then they created a sale price and estimated price maps, per zone. The resulting maps can be viewed in the last section of this chapter.

In Whiterell (2007), a valuation model for single family residential is developed. As explanatory variables, the author constructs 8 neighbourhood clusters: access to railroad, age of the property (there are 5 types of age classifications), some property characteristics like fireplaces, and land size classifications. The author finds an interesting linear relation between estimated prices and real prices.
With the objective to minimize the spatial correlativity, the model uses clustered census tracts; the cluster analysis is an exploratory data analysis to organize information about variables so that relatively homogenous groups can be formed.

3.2 Location Variables

Determining and developing location variables for real estate valuation is one of the main objectives of this thesis. The objective of this chapter is to review some of the location variables in literature and from there, later in the thesis, define and develop the ones to use in the model.

The importance of location has been long recognized in property valuation, but before the development of GIS, it was difficult to accurately measure the impact of location in real estate models (Rodriguez et al. 1995). In fact, Wyatt (1996) mentions the influence of location on property value, as the most important factor. For instance, Bitter et al (2007) found evidence that the marginal price of important housing characteristics varies over space. Basu and Thibodeau (1998) state that location characteristics, such as distance to major transportation arteries, quality of public schools, crime rates, proximity to nonconforming land use, etc. are more difficult to measure than structural characteristics and are rarely included in the available data; but regardless of the difficulties, their impact is important for hedonic house price estimation.

There are various ways to measure location, one of which is accessibility. Wyatt (1997) describes accessibility as an aggregate measure of how reachable a location is from other locations. This is a relative measure resulting in comparing locations of desired facilities.

The impact of the location facilities can be measured by location variables, which can have different connotations and levels. The traditional, and probably the most common location variables, are related to the proximity to the Central Business District (CBD). The traditional monocentric models assume that employment is at a single point in space, called the Central Business District (CBD). A household utility is assumed to be dependent on the distance or accessibility of its house to the city center and to the other goods (Borst 2006). There are a number of critiques to the traditional monocentric models; the main critiques are that modern urban areas are not centralized, they are decentralized in both commercial and labour localization (Branas-Garza et al. 2002), and a number of sub-CBD are common in modern urban areas.

In modern models of urban valuation, proximity to CBD is an important variable, but there are other important variables that are now been considered. In mass appraisals, two levels of location variables can be identified. First, the attributes that are defined by their proximity to the property, for instance distance to CBD, distance to shopping, distance to education facilities, etc. Second, the location attributes that characterize the neighbourhood, like prestige or reputation of the neighbourhood, average income of the households within the neighbourhood, racial and ethnic characteristics, etc. In a way it can be stated that the former are more rational, and the latter attributes are more subjective. As Chesire and Sheppard (2004) state, it might be more dependent on the ability of a household to choose a neighbourhood within its desired characteristics.

An interesting review of the use of GIS for defining location variables in valuation models is done by Rodriguez et al. (1995). The authors show how GIS can provide location variables to the traditional valuation models. They estimate a model using residential sale prices, property characteristics like size and age, and location variables represented by distance to CBD. They compare the results of a model using straight line distance and a model using a distance calculated from a shortest path algorithm. The relevance was similar, but they indicate that the value of the shortest path was more significant. This study is an example of the use of the traditional distance to CBD variable with different approaches to measure this distance, but is not the only location variable used in the literature.

There are quite a few location variables used in the literature, and probably the most complete review of location variables in real estate is done by Kryvobokov (2007). The author identifies the most important location attributes for market value by analyzing different studies. To estimate the most important location attributes, the process is divided in two stages: extraction of attributes and estimation of their weights. The author identifies two types of statistical models commonly used, which are traditional linear regression models and logarithmic regression models.

For instance, in the case of the linear regression models, he uses a weight equation to compare the importance of each of the location variables. Being β_i the estimated parameter of the location variable, the weight is calculated as:

$$W_i = \frac{|\beta_i|}{\sum_{i=1}^{n}|\beta_i|} \qquad (3\text{-}1)$$

The sum of the weights has to be 1.

The author identifies several location variables and states the difficulty in comparing them among different studies, mainly because the variables are used in different regression approaches, with different units and for different objectives. He also identifies that in developing countries the lack of data shortens the number of variables used in the studies.

In conclusion, most of the models use, as dependent variables, the sale price, marked-based assessed value, or market rent. As explanatory variables, the most common location variables are:

- CBD accessibility
- Demographic characteristics
- Water objects accessibility
- Green area accessibility
- Income level of population and prestige
- Commercial objects accessibility and characteristics
- Road accessibility and characteristics
- Planning and urban developments characteristics
- Nuisance proximity
- Education level of population
- Crime level

The variables with the highest average weight in the studies are: proximity to CBD, proximity to commercial facilities, access to education facilities, prestige (measured sometimes with income), and crime.

Having the above study as a starting point, other studies are presented, in which many authors have analyzed the importance of location variables and the externalities (negative or positive) they cause in property prices, housing demand, and city configuration.

3.2.1 Distance to Facilities

Branas-Garza et al. (2002) analyze the differences in prices of housing in terms of distances to CBD. They base their study on three main questions, and the most relevant is: Is the distance to the CBD really a determinant? They find, as expected, that the variable that represents the classical distance to the CBD has a negative impact, the greater the distance the lower the price, but with low elasticity, meaning that it does not provoke high variations in prices. In the study they also find that the quality of the neighbourhood is relevant, but also with a low elasticity. The authors conclude that there is no homogeneity in the housing supply, and distances are not significantly relevant for the determination of house prices. Despite how high the house prices are in the central area, there are high price spots outside too.

Bowes and Ihlanfeldt (2001) study the effect of proximity to rail stations in property prices. They identify four factors in station areas, two which may increase the value of the property, and two which may lower the value. The positive factors are, first, the accessibility advantage of having close rail stations might reduce commuting time and savings in driving stress. Second, the commercial services that are usually close to stations increase the commercial area of the surrounding neighbourhoods. As negative factors, the first is the externalities emitted by the station like noise, pollution, and the ugly view of a rail station. The second negative factor is that criminality can increase, because a station increases the accessibility of strangers to the neighbourhood. In their model they support this idea of higher crime rates near the station. Retail generated by the rail station is shown to be important when owners are far from the CBD. The benefits of less commuting time are higher than crime or retail effects. Retail effects are larger than crime effects, but not when the neighbourhood is close to the downtown area. Large positive effects are found especially in high income neighbourhoods, where they are one quarter and three miles in proximity to the station. Negative crime effects are found mainly close to downtown, especially when the station has parking.

Des Rosiers et al. (1996), study the effect of both proximity and size of shopping centres on the value of residential properties. Similar to other sources of externalities, proximity to shopping centres simultaneously generates attraction, as well as repulsion effects, in the household's decision of location. The negative externality is the result of noise, congestion, and pollution from being in close proximity to the shopping centre. The positive is the convenience of having easy access to shopping and especially the savings in travel costs. They found a clear positive impact on the size of the shopping centre and residential value. Calculating an ideal distance, too close generates some negative externalities; the optimal distances are: 215 m for a neighbourhood size shopping centre, 310 m for a community size shopping centre, and 532 m for a regional shopping centre.

In a second study, Des Rosiers et al. (2000), measure the effect of proximity to primary schools on surrounding residential values, using hedonics. Proximity to schools has two effects on the household's decision of location. On one hand, there is a negative externality as a result of noise, traffic, or even vandalism; while on the other hand there is a positive externality resulting in less travel time. In the end, the authors conclude that easy access to schools is still is more important than the negative proximity effects. They have found an optimal distance to be about a 12 minute walk from home.

Yang (2001) in the context of Beijing, China, takes as location variables the traditional distance to CBD, and four dummy variables, three related to where the property is located, west, north, or south, and one indicating if the property has access to public facilities. Public facilities are those existing in the residential district, for instance, education facilities (3 types), recreation centres, clinics, grocery stores, and a police station. If more than four are provided, the area acquires the value for facility of one. The study determines that the most important location factors are the directional ones, which are highly positive for west and north, and strongly negative for south. The dummy facility is also strong but negative, which is strange, because logic says that the presence of facilities will increase the price. The distance to CBD is negative, which is consistent with the logic that a longer distance to CBD results in a lesser price. The explanation the author gives, for the negative value of the facility variable, is that in China, developers have to pay high management fees and taxes because of the public facilities to local government, which in some way may decrease the demand for some properties.

In Chesire and Sheppard (2004), the authors analyze the impact of proximity to public goods in house prices. For example, the value of an average house can be increased by 30% for having easy access to good schools. Other important location factors used in this study are distance to CBD, percentage of neighbourhood in industry, and racial composition factors.

3.2.2 Neighbourhood Characteristics

3.2.2.1 Racial and Ethnicity

The racial factor in real estate value has been largely discussed in many studies. For instance, Clapp and Ross (2004) examine the evolution of housing prices related to the schools performance and student racial and ethnic composition. They find that there is been a declining tendency, over time, for racial and ethnic school segregation and that there is no evidence that ethnic changes influence housing prices. They mention that segregation in Conneticut is attributable to sorting based preferences of the different groups. This might be in relation to the price that they did not fully explain, and it is contradictory to other studies that show strong racial segregation, and negative influence in housing prices because of racial differences (Thaler 1978; Chambers 1992; Chesire and Sheppard 2004). Racial or ethnicity does not seem to be a big factor in Bogota, since the population is highly homogeneous, with no clear racial or ethnical differences.

3.2.2.2 Crime

It has been extensively studied in terms of spatial distribution of crimes, and psychological reasons. Urban crime has also been studied, and some authors affirm that there is a strong positive relation between size of a city and crime, but in cities, crime is spatially concentrated (Gibbons 2004).

In the context of valuation models, the first highly cited study was conducted by Thaler (1978) in Rochester. He estimates a hedonic model taking in account crime rates per census tract. For data input, different types of crimes were obtained: total offences, property crimes, crimes against persons, and property crimes committed in or around homes. Since these variables are all highly correlated, the author focuses on property crimes, because they are crimes with a more accurate location. The study finds a negative relationship between prices and crimes; an increase of one standard deviation in crime means a 3% decrease in the average price of the sample.

Gibbons (2004) states that crime has a significant negative impact on prices. Urban crime is a powerful influence on the perception of area deprivation. Fear of robbery, burglary, and theft promotes insecurity and nervousness. The spatial concentration of crime can have dynamic effects driven by household location decisions; neighbourhoods can enter into a spiral decline, areas with high crime and unemployment rates acquire a bad reputation, and consequently, people, shops, and places of employment leave. However, crime itself is not a price determinant; it is the type of crime that is important. For instance, crimes like vandalism and graffiti have a significant negative impact on prices. In contrast, the study does not find a direct impact of burglaries on prices. It is surprising that prices respond more to acts of criminal damage than to burglaries given the apparent physical and emotional cost of the latter. The explanation offered by the Gibbons (2004) is that vandalism and graffiti are important factors that motivate fear of crime in the community, even though the evidence in the study suggests that these types of crimes are not strongly correlated with more serious crimes. The other explanation is that graffiti and vandalism can be symptoms of community instability and neighbourhood deterioration.

Lynch and Rasmussen (2001) discuss the difficulty of establishing the seriousness of the crimes, and how this measure can be relative to different urban contexts. They do not find a significant reduction in the average house price due to crimes in their location in Jacksonville, Florida. Although, they state that crime does not substantially affect the price of the average home, in high crime areas house values decline dramatically. Neighbourhoods characterized as high crime zones show a negative relation between property crime rates and prices.

Concluding the crime studies, it is possible to state that not only do crime rates have an impact on the price of properties, but also, and most importantly, it impacts the reputation of the neighbourhood. Vandalism has more impact than actual crimes, but there is direct relation between the zones with high crime rates and zones with a lower perception of security. Lynch and Rasmussen (2001) mention that a crime in a good reputation neighbourhood might not have the same impact on the price of the properties as the same crime in a neighbourhood with a bad reputation.

3.2.3 Several Location Variables

Yavas et al. (2003), in the context of Istanbul, study the rents of 817 properties. As location factors they include availability of transportation facilities, green areas, view, and shopping facilities. The methodology used in this study is different than the traditional approaches to measure the location variables. They measure the location variables as levels of satisfaction with the availability of any of the variables. Another location variable they used is where in the city, the properties are located, finding a positive impact with the properties located in the European part of the city. The other location variable with a negative relation is view, while all of the other variables have a significant positive weight in the model, meaning that a higher level of satisfaction equates a higher price.

An approach that does not involve regression, and is used mainly because the market data is not available or unreliable, is applied in Kryvobokov (2005) study. In the context of the city of Donetsk, in Ukraine, the author estimates weights for the location attributes by using a pair-wise comparison survey to experts that have a good knowledge of the local real estate market, as well as the neighbourhoods. He selected 20 experts from four groups: appraisers, realtors, urban planners, and land managers. The experts had to compare and respond to which location attribute they considered more important from a list of presented cases. In other words, they had to choose the best situation between two location cases; for instance, 1 km closest to the CBD versus an area with more crime rates.

The results are analyzed using an Analytic Hierarchy Process (AHP) and direct comparison. For the experts in the groups of land managers, realtors, and valuers the highest weight is for prestige (wealthier neighbourhoods), while for the urban planners it is crime, CBD, and local CBDs.

In Kryvobokov and Whilhelmsson (2007), a model to determine statistically significant location attributes is developed. In the context of the city of Donetsk, Ukraine, the sample size is 325 apartments, using the following location attributes:

– The accessibility to the CBD; distance in Km
– The accessibility to the nearest secondary centre; distance in Km
– The accessibility to the nearest public transportation stop; distance in Km
– The accessibility to the railway station; distance in Km
– The accessibility to water (river, ponds, etc.) and green areas; distance in Km.
– Nuisance proximity; distance in km
– Prestige; dummy for prestigious area.

As CBD the authors consider the centre of the city, and 14 secondary CBD areas with relative high concentrations of business and commercial development. Comparing the importance of the weights, they identify in the first model, (linear model taking into account all the apartments) distance to nuisance to be the strongest (positive), distance to water and green the second (negative), and distance to CBD the third (negative); all of the others are statistically non significant. In the second model, they compare (linear model taking into account only the apartments outside the central administrative district) distance to water and green areas first, distance to local CBD second, and distance to CBD third, all of which are negative.

In Lake et al. (2000), the objective of the study is to quantify how different land uses would affect house prices, taking into account what is visible from a property, in the context of Glasgow, Scotland. Four groups of variables are used; first, structural variables, direct characteristics of the property, second, neighbourhood variables, variables that describe the quality of its surroundings, third, accessibility variables, calculated as the Euclidean distance from each property to shops, parks, railways, schools, bus routes, and the CBD, and fourth, environmental characteristics, like noise and indicators of the type and extent of land uses, which can be seen from the property. To measure the visual impact, the authors identified land visible from each property by creating a Digital Elevation Model (DEM). The methodology described above produced 327 possible explanatory variables for each property. However, the authors do not include all the explanatory variables in the regression because many of the variables are highly collinear. In conclusion, they indicate that the variables derived using GIS have been successful in accounting for the variety of factors which might have an impact in property prices.

Pompe and Rinehart (1995), use a hedonic model based on OLS, to study the impact of distance to the nearest beach and the width of the beach on property values. They found a decrease in property price for farther and narrower beaches.

Powe et al. (1997) hypothesis is that woodlands are recreational amenities, such as a park, and that house buyers may pay a premium for easy access to woodlands, as well as for the aesthetic reason of living nearby. They use GIS to measure the distance to forest sites and the area of those forest sites. The study finds benefits associated to the close proximity to woodlands and within easy reach to larger areas. They emphasize that the results apply only to the studied region in the south central UK.

4. DATA

4.1 Fontibon District in Bogota, Colombia: case-study

The study area is one of the 20 districts in Bogota, the Fontibon District. The District itself is the size of a small city, and is located in the west part of Bogota. It has 312,629 inhabitants, representing 4.5% of the total population of the city, and its size is 3.325,88 hectares with 80 neighbourhoods (Secretaria de Gobierno Bogota 2006). Fontibon is a particularly heterogeneous district with interesting elements for this study, such as El Dorado International Airport, access to three of the most important roads in the city, a variety of land uses like office sector, leisure, local commerce as well as residential, and it has important industrial facilities. In addition, it is close to the historical downtown of the city, less than 10 km, and has limits with the most important urban park in the city, Simon Bolivar Central Park.

Figure 4-1 shows the location of Fontibon in Bogota.

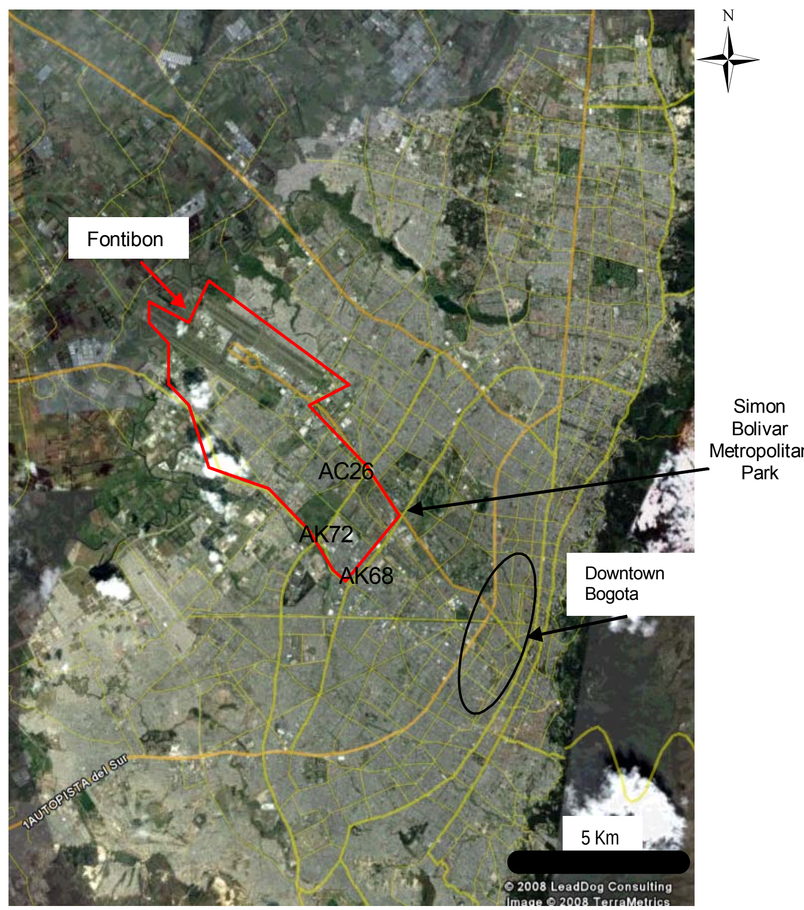

Figure 4-1 Location of Fontibon in Bogota. (from Google Earth)

4.2 Data Collection and Description

After some depuration a total of 23,344 parcels were used for the study. Property attributes, such as age and values, and shape files for the District were provided by the Cadastre of Bogotá. The Observatory of Public Safety of Bogotá (SUIVD) provided the data about crime and road networks. The description of the data and characteristics are shown in the Table 4-1.

Table 4-1 Description of data used for the study

Data	File type	Attributes	Source
Parcels (lots)	Shape file (Polygon)	Area, Perimeter, Parcel Code	A
Border of Fontibon District	Shape file (Polygon)	Area, Perimeter	A
Sub-districts	Shape file (Polygon)	Area, Perimeter, Strate	A
Property Information 1	Database (alpha numeric)	**Area in square metres**, **value per square metre of land**, total value, year of the valuation, address, constructed area, **value per square meter of construction**, code for sub district, a code for neighbourhood, **code for parcel** and code for property.	A
Property Information 2	Database (alpha numeric)	Code district, code neighbourhood, code property, code for use, year of construction, area of use, number of storeys, unitary value, and **land use**.	A
Criminality 1	Shape file (Point)	Burglaries in Fontibon during 2006	B
Main City Roads	Shape file (Polyline)	Roads, names, type of road, and nomenclature	B

Sources

A: Cadastre of Bogota
B: Observatory of Public Safety (SUIVD) of Bogota

4.3 Data Preparation

The objective of the data preparation is to link attributes to geometric data. Specifically, to construct a land use map and a residential parcel map by creating attribute tables with: location, use, characteristics and the value in terms of value per square meter of construction and value per square metre of land.

In terms of land use, 37 land use types can be identified in Fontibon. The main uses are: residential (73% of the properties), commercial (7%) and warehouse (4%).

The parcel shape file is the one that limits the detail of the analysis, and there are a number of properties and buildings for the same polygon or parcel. The assumption done for this study is that the characteristics of each parcel are the average characteristics of the properties or buildings in this parcel. The flow of the analysis is shown in the Figure 4-2.

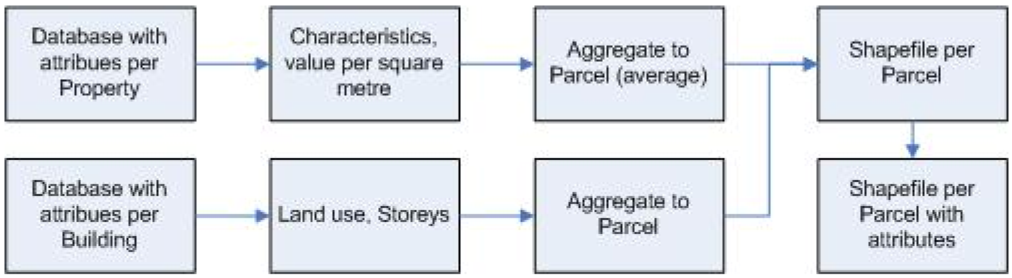

Figure 4-2 Flow chart of the data preparation to obtain the shapefiles with attributes.

The database with values per square metre provided by the Cadastre of Bogota, which will be used for further analysis, are result of on-site valuations and econometric models that take into account characteristics of the property and quality of the construction, among other specific property related attributes. It would be ideal to have a comprehensive database with market values to make a comparison, but such a database is not existent in Bogota; the existent market values data is biased since only contain medium to high standard properties.

5. CREATION OF VALUE MAPS AND LOCATION VARIABLES

5.1 Value Maps

An important application of GIS in real estate valuation is the possibility to create value maps. These maps reflect the location of the properties and its prices, and provide possibility to analyze patterns of prices.

There are two main types of land value maps: discrete and continuous. The discrete maps are the value of each parcel, as shown in Figure 5-1. The continuous maps are surface value maps that use interpolation techniques to estimate a surface value. See Figure 5-2.

5.1.1 Discrete Maps

The first type of value map is a discrete map, showing the average value per square metre of construction for each parcel. The units are in Colombian pesos. See Figure 5-1.

5.1.2 Surface or Continuous Value Maps

The second set of maps is surface value maps (Figure 5-2) constructed by interpolating the discrete values, however, the only objective of performing an interpolation is to help the visualization. These maps are constructed by transforming the parcel polygons to value Point. The interpolation used to create the surface value map is Inverse Distance Weighted (IDW). This is probably the simplest method for interpolation; the process consists in identifying the neighbourhood of the point to be interpolated, and then a weighted average is taken from the observation values in this neighbourhood. Consequently, the estimated values are function of the distance, magnitude, and surrounding points. For the maps presented in this section the IDW was calculated with power 2, number of points 12, and cell size 5 m. This method of interpolation has advantages and disadvantages, but it was chosen for its simplicity. The advantage of this surface map over the discrete map is that it offers a more complete view of the value tendencies over space, offering a good view of the value patterns. From this map we can see that the southeast part of the District seems to be a high value zone.

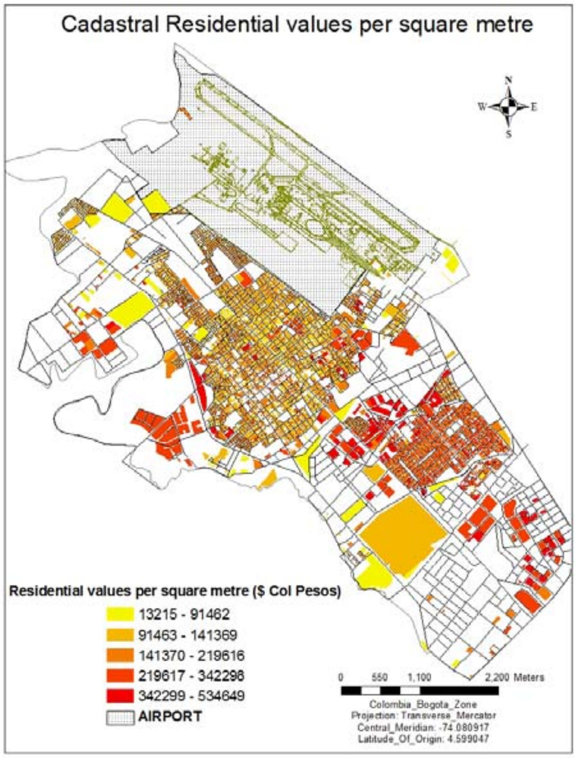

Figure 5-1 Discrete Value Map of the Fontibon District showing the values per square metre of construction per parcel.

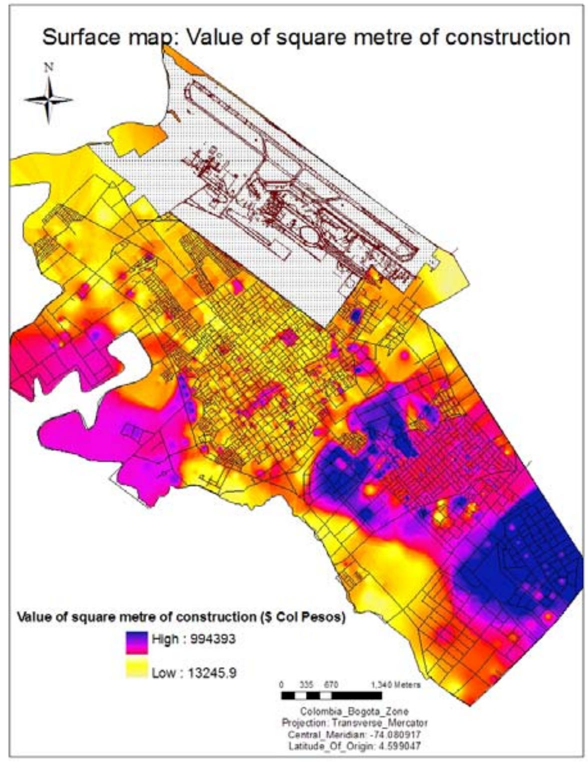

Figure 5-2 Surface Value Map of the Fontibon District showing tendencies of values per square metre of construction

The two type of maps show quasi identical patterns. However, value tendencies are clearer in the surface map Figure 5-2, but the visualization can be affected by errors carried by the interpolation method.

Analyzing the values, the southwest part shows high price areas. It is particularly interesting that there are high constructed prices in the west part of the District. This is usually industrial land, but with new residential developments that might have good construction standards.

5.2 Development of Location Variables and Proximity Attributes

5.2.1 Methods

Location variables are elements that may have an impact in the price of the land and in the price of properties; they are sources of externalities that can generate positive, as well as negative, externalities. In this chapter, first the location variables are presented, and after, a discussion regarding the possible effect that each location variable can have on the value of the parcels is presented. This discussion is based on visual interpretation of the value patterns shown in the value maps, and how they seem to be affected by the proximity to each location variable. In the next chapter, this analysis is done with statistical approaches to be compared with the visual interpretative approach.

Two types of location variables are developed. First, local level variables such as distance to shopping centres, offices, leisure, commerce and criminality. Second, city level variables such as access to main roads, distance to Simon Bolivar Central Park, and distance to the airport. For each variable proximity attributes are designed to capture the impact in the value.

Three main approaches are used for the development of the proximity attributes, as shown in Figure 5-3.

1. Euclidean distance: The shortest straight distance between each property and the required facility. (Network distance could have been used, but it was not used for computational reasons in order to simplify the study).
2. Buffer zones: is used when the number of facilities is not punctual, for example, in the case of access to main roads.
3. Density maps: are used when the punctual facilities are too many and the point to point distance is unmanageable.

The polygon shapes that represent the land use were converted to point shapes in the centroid of the polygon.

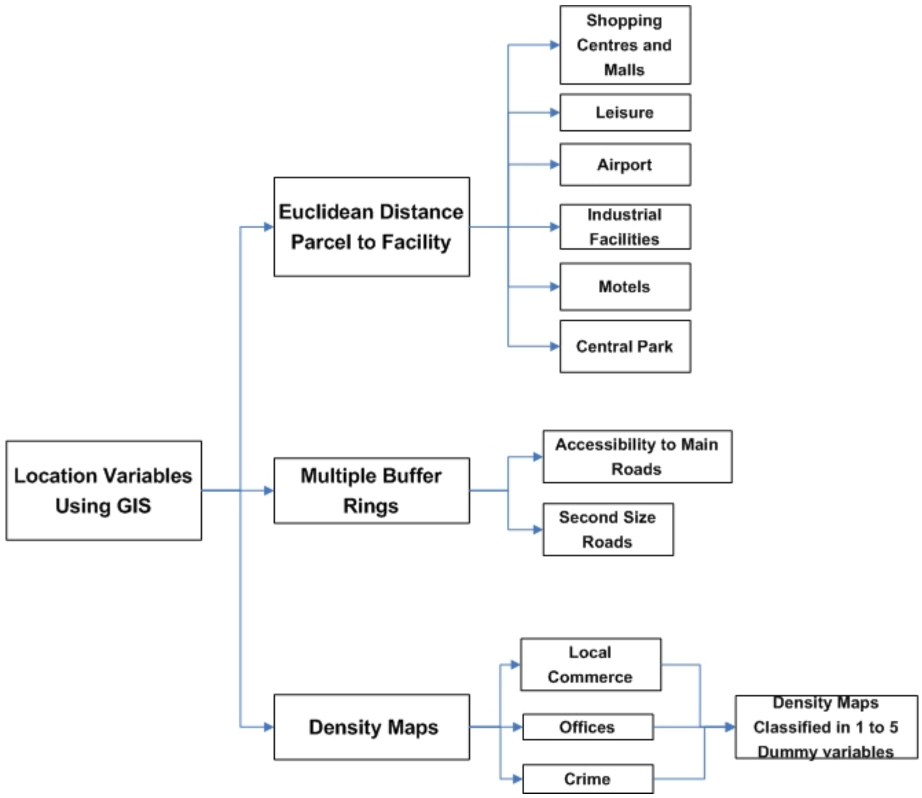

Figure 5-3 Development of Location Variables

5.2.2 Euclidean Distance Parcel to Facility

This method consists in calculating the Euclidean distance from each of the parcels to the objective facility.

The results of this calculation are all the distances from the parcels to facilities, as many distances as facilities. For instance, in the explanatory Figure 5-4, the resulting distances are d1, d2, and d3. For the statistical model, either the minimum or the average distance can be chosen. Five elements are calculated with this method, which are distances to: shopping centres, leisure, Simon Bolivar Metropolitan Park, motels and airport.

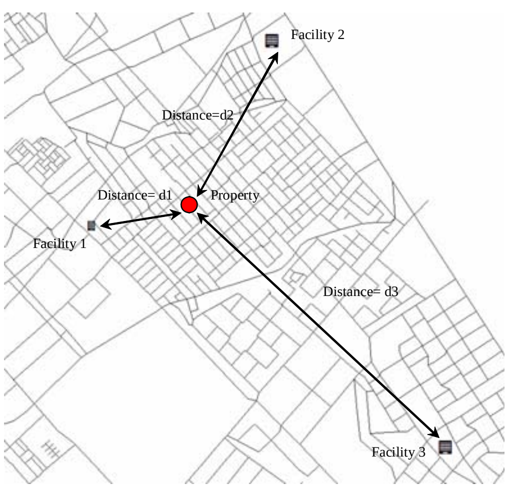

Figure 5-4 Example of Parcel Facility Distances

5.2.2.1 Shopping Centres and Malls

There are two types of amenities in this dataset: Large shopping malls and small shopping centres, both are treated equal. The analysis of distances therefore, is the distance of each residential parcel to each of the shopping facilities

Two variables are taken from this analysis, the average distance of each property to a shopping centre, and the minimum distance of each property to a shopping centre. Both variables will be modeled in the next chapter.

5.2.2.2 Leisure

The leisure facility dataset includes six types of leisure facilities: amusement parks, churches, cinemas, hotels, pools, and restaurants.

For the creation of the leisure variable, only five types are taken into account. Churches are not considered, because they are spread all over the District. As in the case of shopping centres, two distances are going to be calculated, average distance and minimum distance. The average distance offers an important view of the accessibility to the different leisure facilities, and the minimum distance offers the distance to the closest facility, which can be any of the five. In this particular case, the facilities can have different purposes, for instance hotels and cinemas.

5.2.2.3 Industrial Facilities

Fontibon has important industrial zones, especially concentrated in the west side of the District. Figure 5-5 shows the location of the large industries in the District. It is hypothesized that industry might have negative effects on the properties, such as pollution, noise, and ugly view. Nevertheless, it may also have positive impacts since industries are employment centres. The variable is the distance from each parcel to each industry. Average distance to industries and distance to the closest industry will be taken as variables.

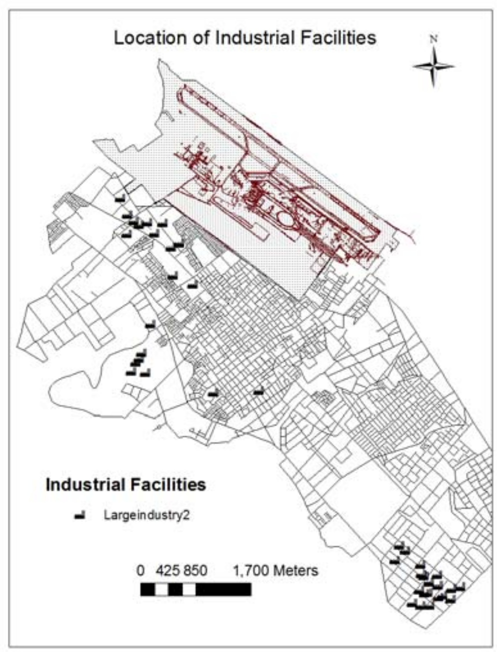

Figure 5-5 Location of industrial facilities in Fontibon, it can be seen that they are mainly in the west side of the District.

5.2.2.4 Motels

This District has several motels, and the hypothesis is that it is a highly negative variable in the context of Bogota. To the author's knowledge, nobody in Bogota wants to live close to this type of parcel use. The distance calculation is similar to the previous variables. The only distance considered for modelling purposes is the distance to the closest motel.

5.2.2.5 Urban Park

Simon Bolivar Metropolitan Park is the largest metropolitan park in the city, and probably largest urban park in all of the country, with an extension of 360 hectares. It is mainly a contemplative natural park, with nice landscapes and a lake. It also has an event square with the capacity to hold over 60.000 people (Bogota Turismo 2008). It does not belong to the Fontibon District, but is in the southeast frontier. Consequently, it will be included in this study. It is envisioned that the park has a positive impact on the price of the properties. For the calculations, the park is represented as a point, and the Euclidean distance from the park to each residential parcel is calculated.

The Simon Bolivar Park variable might have some correlation with the leisure one, and also some spatial correlation with the main roads variable. This will be discussed in the next chapter. The location of the park is shown in the Figure 5-6.

5.2.2.6 Airport

El Dorado International Airport is in the heart of Bogota, and is located in the Fontibon District (Figure 5-6). This airport is the only commercial airport in Bogota and it is just 12 km to the city's downtown area. It is the most important airport in Colombia with over 9 million passengers a year. Some sources put El Dorado as the first airport in Latin America in terms of cargo movements, and the fourth airport in terms of passengers.

It is hypothesized that the noise of the aircrafts is a notorious disadvantage of the nearby properties that is not compensated by easy access to the airport. The calculation of distances is the same as with the previous variables, Euclidian distance from the airport to each parcel.

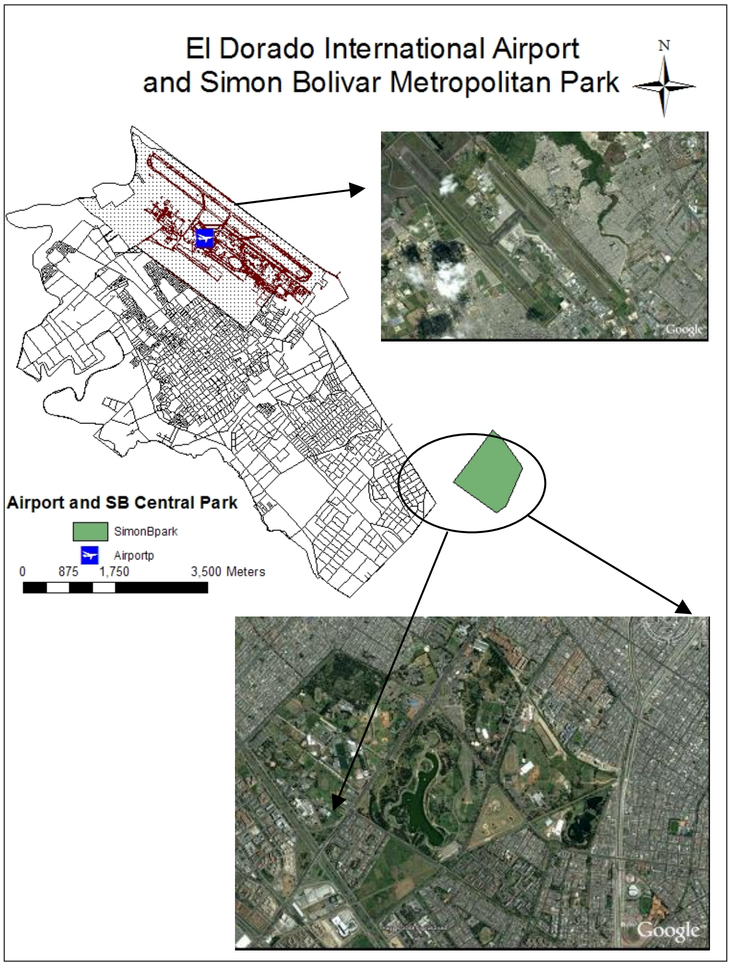

Figure 5-6 Location of Airport and Simon Bolivar Park with respect to the Fontibon District.

5.2.3 Multiple Buffer Rings

This method is used to measure accessibility to roads. It consists of creating rings of distance surrounding the infrastructure, the first two buffers are every 50m, and after, all the rings are every 100 metres. The distance value of the ring is assigned to parcels within the buffer, see Figure 5-7.

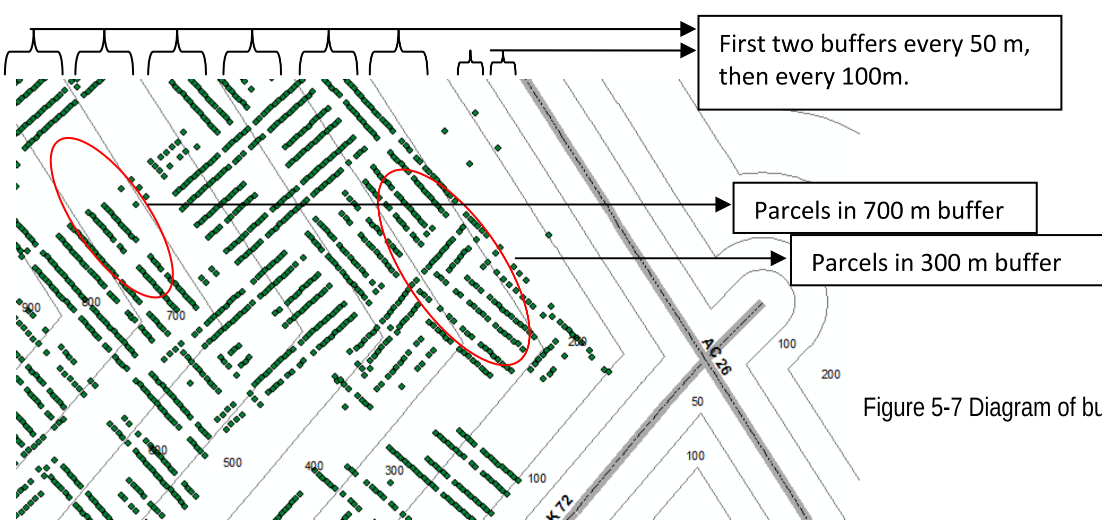

Figure 5-7 Diagram of buffer approach.

5.2.3.1 Accessibility to City Main Roads (City arteries with at least four lanes each direction)

The city arteries are high capacity avenues (4 lanes each direction) that connect the city from west to east or from south to north. They are among the most important roads in the city. It is hypothesized that its proximity does have an overall positive impact because they offer easy access to the rest of the city (e.g. using the AK26, 20 minutes driving time to the downtown). However, their impact might also be negative because of noise and pollution. The avenues selected are AK 26, also called Avenida el Dorado, AK 72, also called Avenida Boyaca, and AK68, also called Avenida del Congreso Eucaristico.

5.2.3.2 Secondary but Important Roads

These are medium-high capacity roads, not just local roads, and are usually considered important roads in the city but not at the top. They offer good connections with the rest of the city, especially at local level, and are extremely important. This variable might be correlated with the variable of main roads. Their effect in the price is hypothesized to be positive, since they increase accessibility. However, they can have some negative externalities.

5.2.4 Density Maps

Density maps are useful in cases where there are too many events, and spatial patterns of behaviour are necessary. Boba (2005) explains that density mapping analysis uses point data to shade surfaces according to the concentration of incidents in particular areas.

For the creation of density maps, usually a ratio is chosen, then the software, in this case the Spatial Analyst, searches for the number of incidents within the radius, and divides the count by the size of the area. Since the values resulting from the density maps are not standardized values, these values are reclassified in five categories of density (5 representing high density, 1 low density). Then the classified values will be used as dummy variables in the model. Therefore, each dummy variable indicates a level of density in a determined area.

5.2.4.1 Local Commerce

A point density analysis with a radius of 300 m is performed. The objective is to determine the areas with more dense commerce. As explained above, the results are reclassified in 5 levels, where level 5 is the higher density. See Figure 5-8 for an example.

5.2.4.2 Offices

This is a similar case to commerce. A point density map is created with a radius of 300 m, and then a reclassification. The density of offices, especially in the southwest side of the locality, seems to be correlated with the presence of industry.

These two elements, offices and local commerce, somehow resemble local CBDs, since they are employment centres and small business districts at local level.

5.2.4.3 Criminality

There are various ways to analyze the spatial distribution of criminality. It is a highly studied science that uses advanced spatial techniques. The traditional mapping and analysis techniques for crime are presented in Boba (2005); a collection of advanced computer applications to crime mapping are collected by Block et al. (1995). A comprehensive review of the main spatial analysis techniques is presented in Anselin et al. (2000). One of the methodologies explained by Anselin et al. (2000) is the Kernel Estimation. The authors explain the Kernel Estimation as a method for examining trends in point data. The idea is to estimate how event levels vary continuously across a study area based on an observed point pattern by creating a smooth map of values. This smoothed map is a density map, and the level of location in the map reflects the point pattern intensity for the surrounding area.

This approach is used in this study to analyze the criminal incidents. The crime points are spread all over the locality. The effect of crime on the price of the properties is difficult to determine because it might have various problems. First, the location might not be exact, unless when there are burglaries to residences. Second, street burglaries do not tend to have a clear impact in the value of the properties. It would be ideal to have information about vandalism. It seems that the effect in the price is in the way of how crimes affect the reputation of the neighbourhood.

The density map of crimes, Figure 5-9, is classified in 5 densities, where 5 has the higher number of crimes, and the dummy variables created are explained at the beginning of this section.

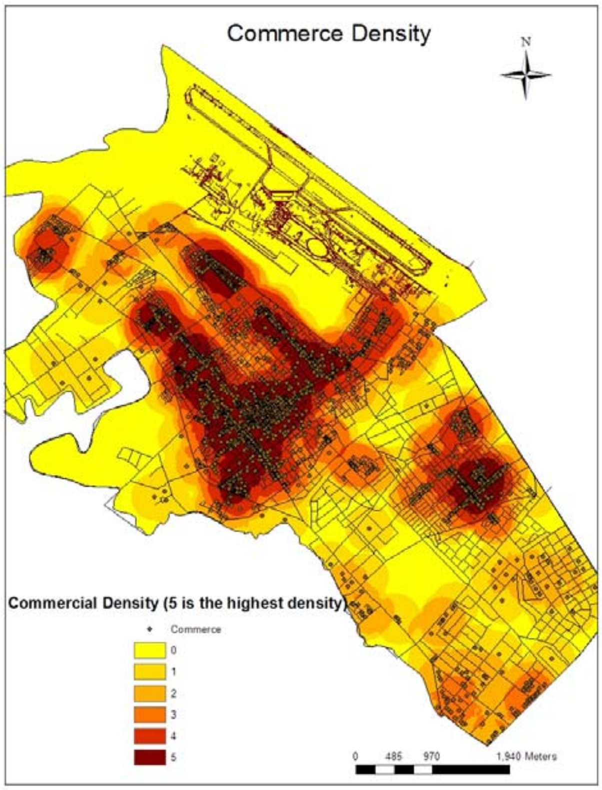

Figure 5-8 Example of a point density map for local commerce

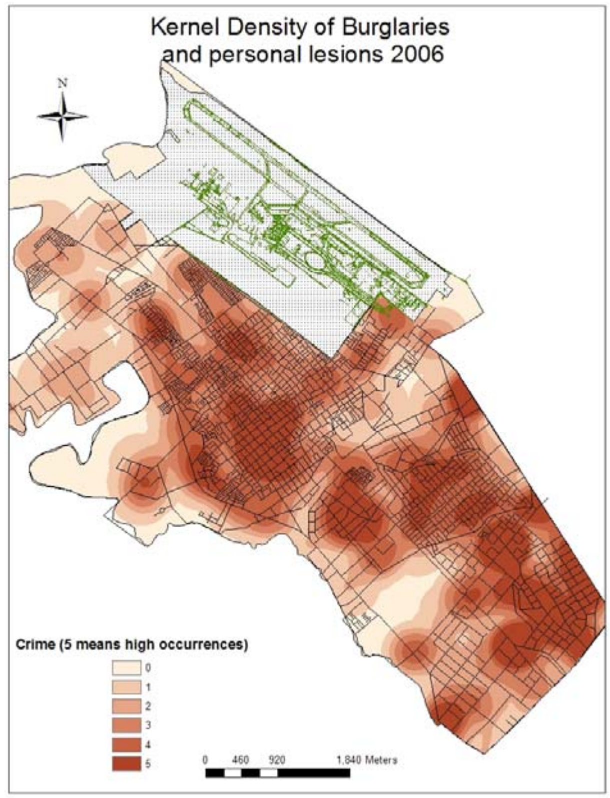

Figure 5-9 Kernel density map for crime (Burglaries and personal lesions)

5.3 Visual Interpretation Analysis

The visual value maps presented are helpful to visualize the pattern of prices, and can be an important instrument to identify the impact of location factors. This has been done in the literature, for instance Kryvobokov and Wilhelmsson (2007), show that the most expensive properties are surrounding the CBDs; or as presented in the value map by Gloudemans and Montgomery (2006), the location of roads was clearly an important factor.

The surface maps are more visually effective than the discrete ones because they offer the information in a direct way, with clear patterns, but also have their limitations, especially in low density sectors. The visual analysis performed in this part is based on the surface maps of land and values of square metre of construction, which both follow similar patterns.

The values of square metre of construction in the databases are usually set by professional appraisers, where the quality of a construction, quantity of square meters and other attributes are taken into account. However, the value maps suggest a relationship between location variables and property values. There are three elements that appear to have a strong impact in the value maps, these are: distance to the airport, accessibility to main roads, and distance to Simon Bolivar Park.

High-priced buildings are favourably located in terms of road accessibility. This might mean that there is a tendency to build better buildings on the sites closest to the main roads. This also applies to the buildings close to Simon Bolivar Park. Nevertheless, it might be correlation between main roads and the park, since the park is strategically located in the corner of two of the main roads. This has to be checked with statistical methods in the next chapter. A new variable will be created, called strategic location, merging these two specific variables into one.

It is also possible that the value is a function of the accessibility to main roads. More precisely, given two buildings of equal quality, the value per square metre of a building located in a highly accessible area is considerably higher than the other building with less accessibility.

Proximity to the airport generates negative externalities, and the value map shows that the farther it is to the airport the more expensive the value per square metre. The same applies to value of the lands; the lands close to the airport seem to be less expensive.

The local amenities do not have an impact as strong as the city factors from the visual interpretation. The one that seems to have a strong effect is the shopping centres; by analyzing the location, they are located in zones of high land and construction value. It seems that there is a correlation between shopping centres and the more expensive areas.

Motels seem to have a negative impact on the price; the location of most of the motels coincides with zones of low value.

There is not a clear spatial tendency between the prices and the location of the industries; they are spread all over the west part of the locality. It can be seen that there are not industries in the most expensive areas, but they are close. However, some expensive areas are in industry zones. There are a number of industries in the north part of the locality, which is one of the zones with less value. The land value maps show that some expensive lands are ones where industries are located; this might be because this is highly productive land, even though in terms of residential parcels it is neither dense nor attractive, it is attractive for industrial use. The overall effect in residential value of closeness to industry is difficult to establish from the visual results.

The secondary roads do not show a clear spatial pattern. Mainly because they strategically cover the entire District, they connect the District from one side to the other. They are clearly in the densest parts of the District, but their effect on the price from the visual perspective is difficult to hypothesize.

Commerce does not seem to have a clear effect. It is mainly concentrated in places with higher population densities. The impact of the location of offices is also difficult to determine. Some of the offices have similar location to the industries but the impact on price of high dense office zones is not clear.

As studied in the literature, crimes as punctual burglaries do not appear to have a direct impact on the value of the land or in the value of the properties. In this case, the number of burglaries does not seem to have a correlation with the value. There is a strong visual correlation between population densities and high-rise buildings with crime events. There are several crimes in the most expensive areas. This might be because burglars prefer to steal in these zones than in the poor areas. The hypothesis that burglary events itself are not determinant of value is mentioned by various authors (Lynch and Rasmussen 2001) (Gibbons 2004) and (Thaler 1978), and appears to be valid in this study. As Lynch and Rasmussen (2001) mention the perception of the people is more determinant of the value of a property, as well as the reputation of the neighbourhood, than number of burglaries.

From the surface value maps it can be hypothesized that the proximity to specific facilities is a variable that determines both the value of construction and the value of land. The visual conclusions from the surface value maps are just at the level of hypothesis. Nevertheless, they offer an interesting hypothesis to be compared to the statistical ones. The econometric models have to confirm the assumption that the real estate value depends on the location, and the importance of each variable.

A summary of the hypothesized impact of the different variables is shown in the table below.

Table 5-1 Summary of Hypothesized Effect of Location Variables

Element	Variables	Effect of Proximity in Price
Shopping centres and malls	Shopave (distance m) Shopmin (distance m)	+
Leisure	Leisave (distance m) Leismin (distance m)	-
Industrial	Industrial min (distance m) Industrial ave (distance m)	+
Motels	Motelmin (distance m)	-
Simon Bolivar	SBpark (distance m)	+
Airport	Airport (distance m)	-
Main City Roads	Mainroaddistance (distance m)	+
Secondary Roads	Localroadsdistance (distance m)	+/-
Offices	Dummy for high density	+/-
Local Commerce	Dummy for high density	+/-
Crime	Dummy for high density	+/-

6 RESULTS

Four approaches are reviewed in this chapter, all of them related to Multiple Regression Analysis (MRA). The first is Pearson correlation coefficient to analyze the correlation of each variable with the value of the properties. The second is the traditional MRA. The third is a step-wise procedure, and finally the local statistical method GWR.

The location variables created in the previous chapter that are used for the models are shown in the table below.

Table 6-1 Description of Variables

Variable	Description
VALUE_M2_C	Same variable, value per square metre of construction
VALUE_M2_L	Value per square metre of land
MINMOTEL	Distance to closest motel
SBOLPARK	Distance to the Simon Bolivar Metropolitan Park
AIRPORT	Distance to the airport
Park Roads	Variable that contains the average of distance to the SB Park and the distance to main roads
MININDUSTR	Distance to the closest industry
MAINROADS	Distance to the main roads
LEISAVE	Average distance to a leisure facility
SHOPPAVE	Average distance to a shopping centre
LEISMIN	Distance to the closest leisure facility
SHOPPMIN	Distance to the closest shopping centre
Comm1	Dummy indicating that the property is in a low commerce density class (class 1)
Office1	Dummy indicating that the property is in a low office density class (class 1)
Crim5	Dummy indicating that the property is in a high crime density class (class 5)
Crim45	Dummy indicating that the property is in a high crime density class (class 4 or class 4)
Comm45	Dummy indicating that the property is in a high commerce density class (class 4 or 5)
Office45	Dummy indicating that the property is in a low office density class (class 4 or class 5)
Comm2	Dummy indicating that the property is in a low commerce density class (class 2)
Office2	Dummy indicating that the property is in a low office density class (class 2)
SECONDROADS	Distance to the secondary roads
Crim1	Dummy indicating that the property is in a low crime density class (class 1)
Comm5	Dummy indicating that the property is in a high commerce density class (class 5)
Office5	Dummy indicating that the property is in a high office density class (class 5)
Crim2	Dummy indicating that the property is in a low crime density class (class 2)
AVGINDUSTR	Average distance to industries
YEAR	Year of construction
Crim3	Dummy indicating that the property is in a medium crime density class (class 3)
Comm4	Dummy indicating that the property is in a medium high commerce density class (class 4)
Office4	Dummy indicating that the property is in a medium high office density class (class 4)
Comm3	Dummy indicating that the property is in a medium commerce density class (class 3)
Office3	Dummy indicating that the property is in a medium office density class (class 3)
STOREY	Average number of storey in the properties
AREA_AVE	Average area of construction
Crim4	Dummy indicating that the property is in a medium high crime density class (class 4)

6.1 Pearson Product-Moment Bivariate Correlation Coefficient

From the values of the Pearson coefficient (Table 6-2 and Table 6-3), it can be inferred that the variable with the strongest correlation is the value of land (for construction value) and value of construction (for land value); this is absolutely consistent with real estate theory, in that more expensive constructions are in the most expensive lands. From the other variables, the ones that that have strongest correlation with the value per square metre of construction and land are: minimum distance to a motel in a positive way, meaning that farther the distance the more price is. Then distance to the Simon Bolivar Park in a negative correlation, meaning that closer to the park means a higher price. The farther to the airport means more value. For the variable composed by main roads and the park, properties that are more close to both are more expensive. Closeness to an industry has a positive correlation; more close to an industry is less expensive. Then, proximity to main roads means that properties closer to a main road are more expensive. Properties closer to a leisure facility (average) are more expensive; the same relation applies for average distance to shopping centres, distance to the closest leisure facility, and distance to the closest shopping centre.
From the dummy variables, the lowest and the highest densities are the ones that are more correlated to the value. Farther distances to areas of low density commerce and offices are less expensive. Properties that are farther from high density crime zones have more value.

Overall, the ranking of the Pearson coefficient values is similar for land values and for values of construction, with the difference that the proximity to the Simon Bolivar Park is the strongest in land value.

The significance of the values, t statistics, has over 99% confidence for most of the variables. For the value of construction, only commerce class 3, office class 3 were significant at 95%, and storey, area, and criminality class 4, were not significant. For value of land, year was significant at 97%, and criminality class 4 and area were not significant.

Table 6-2 Pearson Coefficient for Value of Square Metre of Construction (values sorted descending)

Variable	Pearson Correlation	Pearson ABS	Sig. (2-tailed)	N	Effect of closeness/ presence in Price
VALUE_M2_C	1	1		23344	
VALUE_M2_L	0.675	0.675	0.00	23344	+
MINMOTEL	0.501	0.501	0.00	23344	-
SBOLPARK	-0.500	0.500	0.00	23344	+
AIRPORT	0.498	0.498	0.00	23344	-
ParkRoads	-0.462	0.462	0.00	23344	+
MININDUSTR	0.425	0.425	0.00	23344	-
MAINROADS	-0.379	0.379	0.00	23344	+
LEISAVE	-0.345	0.345	0.00	23344	+
SHOPPAVE	-0.344	0.344	0.00	23344	+
LEISMIN	-0.227	0.227	0.00	23344	+
SHOPPMIN	-0.203	0.203	0.00	23344	+
Comm1	0.179	0.179	0.00	23344	-
Office1	0.179	0.179	0.00	23344	-
Crim5	0.148	0.148	0.00	23344	-
Crim45	0.127	0.127	0.00	23344	-
Comm45	-0.126	0.126	0.00	23344	+
Office45	-0.126	0.126	0.00	23344	+
Comm2	0.115	0.115	0.00	23344	-
Office2	0.115	0.115	0.00	23344	-
SECONDROADS	-0.103	0.103	0.00	23344	+
Crim1	-0.098	0.098	0.00	23344	+
Comm5	-0.096	0.096	0.00	23344	+
Office5	-0.096	0.096	0.00	23344	+
Crim2	-0.090	0.090	0.00	23344	+
AVGINDUSTR	-0.059	0.059	0.00	23344	+
YEAR	0.040	0.040	0.00	23344	-
Crim3	-0.023	0.023	0.00	23344	+
Comm4	-0.022	0.022	0.00	23344	+
Office4	-0.022	0.022	0.00	23344	+
Comm3	0.013	0.013	0.05*	23344	-
Office3	0.013	0.013	0.05*	23344	-
STOREY	0.011	0.011	0.09**	23344	-
AREA_AVE	-0.010	0.010	0.13**	23344	+
Crim4	0.005	0.005	0.43**	23344	-

*Significant at 95% of confidence

** Not significant.

Table 6-3 Pearson Correlation Coefficient for Land Values

Variable	Pearson Correlation	Pearson ABS	Sig. (2-tailed)	N	Effect of closeness/ Presence in Price
VALUE_M2_L	1.000	1.000		23344	
VALUE_M2_C	0.675	0.675	0.00	23344	+
SBOLPARK	-0.669	0.669	0.00	23344	+
AIRPORT	0.651	0.651	0.00	23344	-
ParkRoads	-0.618	0.618	0.00	23344	+
MINMOTEL	0.609	0.609	0.00	23344	-
MININDUSTR	0.552	0.552	0.00	23344	-
SHOPPAVE	-0.510	0.510	0.00	23344	+
MAINROADS	-0.508	0.508	0.00	23344	+
LEISAVE	-0.506	0.506	0.00	23344	+
SHOPPMIN	-0.352	0.352	0.00	23344	+
LEISMIN	-0.319	0.319	0.00	23344	+
Crim5	0.236	0.236	0.00	23344	-
Crim45	0.208	0.208	0.00	23344	-
Comm1	0.178	0.178	0.00	23344	-
Office1	0.178	0.178	0.00	23344	-
Comm45	-0.162	0.162	0.00	23344	+
Office45	-0.162	0.162	0.00	23344	+
AVGINDUSTR	-0.150	0.150	0.00	23344	+
Comm2	0.146	0.146	0.00	23344	-
Office2	0.146	0.146	0.00	23344	-
Crim2	-0.130	0.130	0.00	23344	+
STOREY	-0.121	0.121	0.00	23344	+
Crim1	-0.117	0.117	0.00	23344	+
SECONDROADS	-0.090	0.090	0.00	23344	+
Comm4	-0.079	0.079	0.00	23344	+
Office4	-0.079	0.079	0.00	23344	+
Comm5	-0.076	0.076	0.00	23344	+
Office5	-0.076	0.076	0.00	23344	+
Crim3	-0.073	0.073	0.00	23344	+
Comm3	0.040	0.040	0.00	23344	-
Office3	0.040	0.040	0.00	23344	-
YEAR	0.014	0.014	0.03*	23344	-
Crim4	0.014	0.014	0.03*	23344	-
AREA_AVE	0.011	0.011	0.09**	23344	-

* Significant at 97% of confidence

** Not significant

6.2 MRA approach

The objective of the first two models is to identify the colinearity among spatial variables by the use of correlation matrices. Then, a second set of models will be estimated excluding the strongly correlated variables. The coefficients of the regression of the first model are not analyzed because they are strongly biased due to the strong multicolineality of the variables. Therefore, the objective of the first model is to analyze the correlation matrices, and to identify how correlated the variables are.

The first table (Table 6-4) presents the correlation matrix for a model including most of the location variables, and value of square metre of construction as dependent variable. In this estimation, the most notorious correlation is between the airport, main roads, and Simon Bolivar Park. The airport is correlated to main roads and Simon Bolivar Park. Simon Bolivar Park is strongly correlated to main roads, airport, and average distance to Leisure; this is consistent with the idea that Simon Bolivar Park is a leisure activity. The correlation between the Park and the main roads is also expected, since the Park is located in the corner of two of the three main roads. Leisure is also correlated to shopping, this might be for two reasons; one, that shopping activity can be considered a leisure activity, and second, some of the leisure facilities are strategically located close to the shopping centres.

The second table presented, Table (6-5), is the correlation matrices resulting from a model, using most of the location variables as independent and the value of square metre of land as dependent variable; the values are very similar to the resulting table of the previous analysis.

In summary, the correlated results were as expected, some location variables are strongly, which might lead to biased results. Therefore, in order to obtain trustworthy models, some variables cannot be used in the same models, or at least one has to be aware of the problem.

Table 6-4 Correlation Matrix for a model estimated with value per square metre of construction as dependent variable

Dependent Variable: VALUE_M2_C		MININDUSTR	SECONDROAD	Crim45	Office1	Office45	LEISAVE	AIRPORT	MINMOTEL	MAINROADS	SBOLPARK	SHOPPAVE
Correlations	MININDUSTR	1.000	-0.351	-0.008	0.096	-0.389	-0.810	-0.105	-0.483	0.520	0.042	0.782
	SECONDROADS	-0.351	1.000	-0.158	-0.115	0.431	0.307	0.045	0.435	-0.070	0.003	-0.331
	Crim45	-0.008	-0.158	1.000	0.036	-0.210	0.094	0.025	0.095	-0.199	0.146	-0.101
	Office1	0.096	-0.115	0.036	1.000	0.050	-0.015	-0.114	-0.075	0.084	-0.039	0.011
	Office45	-0.389	0.431	-0.210	0.050	1.000	0.357	0.115	0.320	-0.183	0.004	-0.356
	LEISAVE	-0.810	0.307	0.094	-0.015	0.357	1.000	-0.195	0.419	-0.278	-0.229	-0.971
	AIRPORT	-0.105	0.045	0.025	-0.114	0.115	-0.195	1.000	0.306	-0.793	0.894	0.063
	MINMOTEL	-0.483	0.435	0.095	-0.075	0.320	0.419	0.306	1.000	-0.390	0.459	-0.561
	MAINROADS	0.520	-0.070	-0.199	0.084	-0.183	-0.278	-0.793	-0.390	1.000	-0.724	0.360
	SBOLPARK	0.042	0.003	0.146	-0.039	0.004	-0.229	0.894	0.459	-0.724	1.000	0.036
	SHOPPAVE	0.782	-0.331	-0.101	0.011	-0.356	-0.971	0.063	-0.561	0.360	0.036	1.000

Strong Correlation

Table 6-5 Correlation Matrix for a model estimated with value per square metre of land as dependent variable

Dependent Variable: VALUE_M2_L		MININDUSTR	SECONDROADS	Crim45	Office45	LEISAVE	AIRPORT	MINMOTEL	MAINROADS	SBOLPARK	SHOPPAVE
Correlations	MININDUSTR	1.000	-0.344	-0.011	-0.396	-0.812	-0.095	-0.480	0.516	0.046	0.784
	SECONDROADS	-0.344	1.000	-0.155	0.440	0.308	0.033	0.430	-0.061	-0.002	-0.332
	Crim45	-0.011	-0.155	1.000	-0.213	0.094	0.029	0.098	-0.202	0.147	-0.101
	Office45	-0.396	0.440	-0.213	1.000	0.358	0.121	0.325	-0.188	0.006	-0.357
	LEISAVE	-0.812	0.308	0.094	0.358	1.000	-0.198	0.419	-0.277	-0.229	-0.971
	AIRPORT	-0.095	0.033	0.029	0.121	-0.198	1.000	0.300	-0.791	0.896	0.064
	MINMOTEL	-0.480	0.430	0.098	0.325	0.419	0.300	1.000	-0.386	0.458	-0.562
	MAINROADS	0.516	-0.061	-0.202	-0.188	-0.277	-0.791	-0.386	1.000	-0.724	0.361
	SBOLPARK	0.046	-0.002	0.147	0.006	-0.229	0.896	0.458	-0.724	1.000	0.036
	SHOPPAVE	0.784	-0.332	-0.101	-0.357	-0.971	0.064	-0.562	0.361	0.036	1.000

Strong Correlation

As seen in the correlation matrix, some of the most important variables are strongly correlated, which will lead to biased and inaccurate estimations. For that reason, in the models presented below, the strongly correlated variables are excluded.

The variables main roads and Simon Bolivar Park cannot be used in the same model since they are strongly correlated; this is because they share a similar location, and the Park is located in the corner of two of the three most important roads. To overcome this problem, a new variable merging Simon Bolivar Park and the main roads is created (ParkRoads variable). This variable is the average of the two distances values.

The coefficient presented in the models below is the standardized coefficient, which makes the comparison between parameters easier. In the standardized coefficient the variables are standardized to have a variable of 1, and it is usually done to determine the effect of each of the independent variables in cases where the variables have different units, or the resulting values are too high.

Apart from the estimated coefficient of each variable, the result tables present the weight of each variable in the model. The weight of the location variables is calculated with the equation presented by Kryvobokov (2007), equation 3-1.

$$W_i = \frac{|\beta_i|}{\sum_{i=1}^{n} |\beta_i|}$$

The sum of the weights has to be 1.

6.2.1 Models 1 Traditional Global MRA Estimation

In this set of models, two models are presented and compared, one taking a dependent variable value of square metre of construction, and in the other value per square metre of land.

The resulting models are presented in the table below:

Table 6-6 Model 1 MRA value of land and value of construction

Model 1	VALUE_M2_C			VALUE_M2_L		
	R Square adj.	0.36		R Square adj.	0.59	
	Beta	t statistics	Weight	Beta	t statistics	Weight
(Constant)		5.46			6.68	
MININDUSTR	-0.13	-9.98	0.05	-0.01	-0.75*	0.003
AIRPORT	0.23	28.01	0.08	0.27	41.54	0.10
SHOPPMIN	-0.48	-27.03	0.17	-0.64	-45.17	0.23
MINMOTEL	0.38	33.90	0.14	0.49	54.07	0.18
LEISAVE	0.83	25.62	0.30	0.76	29.24	0.28
Year	0.02	3.65		-0.01	-1.73*	
SECONDROADS	-0.06	-9.53	0.02	-0.06	-10.86	0.02
Office45	-0.02	-2.96*	0.01	-0.05	-9.53	0.02
Crim45	0.01	1.68*	0.00	0.03	6.58	0.01
ParkRoads	-0.62	-19.46	0.23	-0.42	-16.41	0.15

* Not significant at 95%

Table 6-7 Correlation Matrix for Land values.

Correlations	ParkRoads	Year	SECONDROADS	Crim45	Office45	AIRPORT	SHOPPMIN	MINMOTEL	MININDUSTR	LEISAVE
ParkRoads	1.00	0.01	0.26	0.01	0.03	0.13	0.38	0.39	0.66	-0.85
Year	0.01	1.00	0.01	-0.01	-0.01	-0.01	0.03	-0.04	0.01	-0.03
SECONDROADS	0.26	0.01	1.00	-0.12	0.43	0.34	0.40	0.20	-0.11	-0.37
Crim45	0.01	-0.01	-0.12	1.00	-0.23	-0.28	0.15	-0.07	0.15	-0.04
Office45	0.03	-0.01	0.43	-0.23	1.00	0.39	0.25	0.07	-0.17	-0.13
AIRPORT	0.13	-0.01	0.34	-0.28	0.39	1.00	0.38	-0.16	-0.18	-0.23
SHOPPMIN	0.38	0.03	0.40	0.15	0.25	0.38	1.00	-0.29	0.07	-0.77
MINMOTEL	0.39	-0.04	0.20	-0.07	0.07	-0.16	-0.29	1.00	-0.02	-0.09
MININDUSTR	0.66	0.01	-0.11	0.15	-0.17	-0.18	0.07	-0.02	1.00	-0.43
LEISAVE	-0.85	-0.03	-0.37	-0.04	-0.13	-0.23	-0.77	-0.09	-0.43	1.00

The explanatory power of the models are not particularly good, especially the construction model, in which only 36% of the variance of the data is explained by the model. This is easily explained because variables that characterize the property are not included, it only includes location variables. Since the land values are less dependent on property specific characteristics, the explanatory power of the land model is quite good, the model explains 59% of the variance. It is important to point out that the objective of this study is to determine and compare the impact of different location variables rather than fully explain the price of the properties.

The values of the coefficients and the impact of each location variable are similar for both models. The coefficient for proximity to industry, MININDUSTR, is negative, which means that closer to an Industry is more expensive buildings and lands; this coefficient also has the lower weight with respect to the other location variables, meaning that its impact is not transcendental. The sign of the coefficient might not be completely odd, since despite the fact that the industries are spread all over the west part of the District, most of the expensive areas are close to industrial areas; this can be seen in the maps, but the explanatory power of this variable does not seem to be strong.

The airport variable has a positive sign, meaning that the farther from the airport the more expensive the value. This is consistent with the initial hypothesis, but the weight of the variable is low, which is not consistent with the hypothesis because a strong negative effect on residential prices in proximity to the airport was expected.
The proximity to shopping has a negative sign, meaning that the farther to a shopping centre the less expensive are properties and lands, with a strong weight. This is consistent with the initial idea that more expensive properties are close to shopping facilities.

Proximity to motels has one of the higher weights, and indicates that the farther to a motel are more expensive properties, which is consistent with the initial hypothesis.
The most bizarre value is the one for leisure proximity because it has the strongest weight and positive sign, meaning that the farther to leisure facilities are more expensive properties. This is neither consistent with the initial hypothesis nor with the common sense of real estate.

The results presented have to be analyzed with prudence, since they are biased due to a correlation between some of the independent variables. As it can be seen in the correlation matrix, leisure is an extremely correlated variable; it has a strong correlation with the shopping proximity variable and with the park roads variable. There can be several reasons for this correlation. Shopping and frequenting parks can be seen as leisure activities. In terms of correlation, due to spatial location the main leisure facilities are easily accessed from the main roads; this also applies for shopping centres. These strong correlations might result in biased estimations of the parameters.

To try to obtain a better view of the importance of each variable in the next set of models, a stepwise procedure is applied to help identify the most significant price determinants amid the proximity variables.

6.2.2 Models 2 step-wise Procedure

With the stepwise procedure, the variables that do not have explanatory power are excluded, estimating only the statistically important variables. This procedure does not completely fix the problem of correlation of the spatial variables, but it helps.

A total of four estimations are presented, two for land value and two for the value of construction.

6.2.3 Models for Land Value

These models are estimated by taking as a dependent variable the value per square metre of land. With the experiences from previous estimations, it is not wise to put together some of the location variables. For instance, leisure and RoadParks are strongly correlated.

The first model excludes the leisure proximity variable, so the RoadPark variable and six more variables are included.

Table 6-8 Stepwise regression 1 for land value

Stepwise Land 1	VALUE M2 of Land			
	R Square Adj	0.55		
Attributes	Beta	t statistics	Significance	Weight
(Constant)		37.43	0.00	
AIRPORT	0.38	62.26	0.00	0.39
MINMOTEL	0.32	46.56	0.00	0.33
ParkRoads	-0.15	-21.88	0.00	0.16
Crim45	0.08	15.35	0.00	0.08
LOCALROADS	0.03	6.14	0.00	0.03
Office45	0.01	2.15	0.03*	0.01

* Not significant at 97%

Table 6-9 Correlation Matrix for Stepwise regression 1 for land value

Correlations	AIRPORT	MINMOTEL	ParkRoads	Crim45	LOCALROADS	Office45
AIRPORT	1.00	-0.11	0.37	-0.35	0.20	0.30
MINMOTEL	-0.11	1.00	0.62	0.19	0.28	0.19
ParkRoads	0.37	0.62	1.00	0.09	0.23	0.16
Crim45	-0.35	0.19	0.09	1.00	-0.13	-0.26
LOCALROADS	0.20	0.28	0.23	-0.13	1.00	0.35
Office45	0.30	0.19	0.16	-0.26	0.35	1.00

For the second model, the ParkRoads variable is not taken in account, and nine other variables are considered.

Table 6-10 Stepwise regression 2 for land value

Stepwise Land 2	VALUE M2 of Land			
	R Square Adj	0.57		
Attributes	Beta	t statistics	Significance	Weight
(Constant)		54.46	0.00	
AIRPORT	0.32	50.16	0.00	0.30
MINMOTEL	0.41	56.37	0.00	0.39
SHOPPMIN	-0.17	-28.31	0.00	0.16
Crim45	0.04	7.84	0.00	0.04
LEISMIN	-0.04	-8.00	0.00	0.04
LOCALROADS	0.01	2.30	0.02	0.01
Office45	-0.02	-4.33	0.00	0.02
MININDUSTR	0.03	4.32	0.00	0.03

Table 6-11 Correlation Matrix for Stepwise regression 2 for land value

Correlations	AIRPORT	MINMOTE	SHOPPMIN	Crim45	LEISMIN	LOCALRO	Office45	MININDUSTR
AIRPORT	1	-0.12	0.26	-0.3	0.1	0.27	0.37	-0.3
MINMOTEL	-0.12	1	-0.21	-0.07	0.11	0.31	0.21	-0.69
SHOPPMIN	0.26	-0.21	1	0.22	-0.5	0.17	0.08	0.13
Crim45	-0.3	-0.07	0.22	1	0.03	-0.15	-0.24	0.22
LEISMIN	0.1	0.11	-0.5	0.03	1	-0.05	0.17	-0.07
LOCALROADS	0.27	0.31	0.17	-0.15	-0.05	1	0.38	-0.31
Office45	0.37	0.21	0.08	-0.24	0.17	0.38	1	-0.21
MININDUSTR	-0.3	-0.69	0.13	0.22	-0.07	-0.31	-0.21	1
	Strong Correlation							

The variables presented in these two models explain fairly well, 55% and 57% of the value of the land. The values for the coefficients are consistent with the visual hypotheses; nevertheless a strong correlation between some variables is still present. The variables with the strongest weight are in proximity to the airport and in proximity to a motel, both with positive signs, meaning there is a negative impact on the price for proximity, which is perfectly consistent with the initial hypothesis and the structure of the city.

Proximity to shopping centres also has a strong weight, but a negative sign, meaning that properties that are closer to shopping centres are more expensive. The other variables do not seem to have a high impact on the price, but it is worthy to mention that high criminality zones appear to be fourth in importance, reducing the value. This is consistent with the logic, but needs a cautious interpretation since, as seen in the literature review; criminality is an extremely complex variable to interpret.

6.2.4 Models for Value of Construction

The models presented in this section are estimated by taking as a dependent variable the square metre of construction.
The fist model only considers proximity variables as dependent variables.

Table 6-12 Stepwise regression 1 for value of construction

Stepwise Construction 1	VALUE M2 of Construction			
	R Square Adj	0.34		
Attributes	Beta	t statistics	Significance	Weight
(Constant)		70.32	0.00	
MINMOTEL	0.36	55.10	0.00	0.47
AIRPORT	0.27	37.94	0.00	0.36
LEISMIN	-0.06	-8.37	0.00	0.07
SHOPPMIN	-0.05	-7.31	0.00	0.07
Crim45	0.02	2.95	0.00	0.02

Table 6-13 Correlation Matrix for Stepwise regression 1 for value of construction

Correlations	MINMOTEL	AIRPORT	LEISMIN	SHOPPMIN	Crim45
MINMOTEL	1.00	-0.56	0.09	-0.21	0.15
AIRPORT	-0.56	1.00	0.05	0.29	-0.20
LEISMIN	0.09	0.05	1.00	-0.51	0.08
SHOPPMIN	-0.21	0.29	-0.51	1.00	0.24
Crim45	0.15	-0.20	0.08	0.24	1.00

The prediction capacity of the variables of this model is 34%. The coefficients of the variables are similar to the previous models; proximity to motels and airports are the variable with the highest weight, both with a positive sign, meaning that farther is more expensive. Proximity to leisure and shopping do not seem to have a strong weight, but their sign is negative, which is consistent with the visual hypothesis. High criminality zones have a positive sign, which means that less expensive properties are the ones located in high crime zones.

A second model is estimated, which differs from the previous models, and considers some specific attributes that characterize the property, such as age of the property, number of storeys, and area of construction. In this model the non-standardized coefficients are also presented.

Table 6-14 Stepwise regression 2 for value of construction

Stepwise Construction 2	VALUE M2 of Construction				
	R Square Adj	0.35			
Attributes	Unstandardized Coefficients	Beta	t statistics	Significance	Weight
(Constant)	85887.49		32.70	0.00	
MINMOTL	30.66	0.32	40.99	0.00	0.34
AIRPORT	17.80	0.30	43.59	0.00	0.32
STOREY	8214.03	0.13	24.30	0.00	0.14
Crim45	5476.52	0.05	8.00	0.00	0.05
AREAAVE	-14.40	-0.04	-8.31	0.00	0.05
ParkRoad	-3.05	-0.07	-7.92	0.00	0.07
Age	-14.20	-0.03	-4.84	0.00	0.03

Table 6-15 Correlation Matrix for Stepwise regression 2 for value of construction

Correlations	MINMOTL	AIRPORT	STOREY	Crim45	AREAVE	ParkRoad	Age
MINMOTL	1.000	-0.209	0.084	0.270	0.075	0.595	0.055
AIRPORT	-0.209	1.000	0.072	-0.285	-0.049	0.328	0.014
STOREY	0.084	0.072	1.000	0.039	-0.123	0.001	-0.011
Crim45	0.270	-0.285	0.039	1.000	0.019	0.152	0.013
AREAVE	0.075	-0.049	-0.123	0.019	1.000	0.081	0.005
ParkRoad	0.595	0.328	0.001	0.152	0.081	1.000	0.052
Age	0.055	0.014	-0.011	0.013	0.005	0.052	1.000
	Strong Correlation						

The R^2 is at a fairy typical level compared to the previous models (35%). The location variables are the ones with higher weights and this is not surprising considering the previous models. The number of storeys appears to be quite an important variable, because properties with more storeys tend to be more expensive. Area of construction has a negative impact, and greater constructed areas have less value per square meter. Age also has a negative impact; the older the property is the less expensive. The tendencies of these two variables agree with real estate sense. The sign for Park is consistent, but the weight seems to be to low. The only odd sign is the one for crime; it is positive which means that the more expensive properties are in high crime zones, which is not completely odd because robbers tend to go to the wealthiest neighbourhoods.

6.2.5 General comments regarding global regressions

One of the main problems with MRA is the colinearity among spatial variables, which leads to biased models, and notably affects the results.

The goodness of fit (R^2) value for the global regressions estimated leaves 60% to 40% of the variance in values unexplained. Some of this unexplained variance is because the model is mainly using location variables, and more variables are required to increase the explanatory power of the model. The low R^2 values are also probably the result of assuming the relationships in the model to be constant over space. The traditional global models, as they are presented above, dominate the literature but are not always appropriate with spatial data.

Apart from the spatial autocorrelation problem, global models ignore potential variations over space; the process that is investigated might not be constant over space, and global models cannot capture the non-stationary component of the relationship (Fotheringham et al. 2002).

For the reasons just mentioned, a local statistical model will be estimated to examine the local spatial behaviour of the models, and analyze whether spatial non-stationarity exists in the main variables.

6.3 Local Statistics Approach: Geographically Weighted Regression

As seen in the literature review section 2.2.2.4, the local statistical models calculate localized parameter estimates and localized versions of all standard regression diagnostics including the goodness of fit R^2. These statistics are estimated for each point and the visualization is ideally using GIS. For the models estimated, the results are presented in surface maps created using IDW interpolation.

The estimations were performed with the Software GWR release 3.0 by (Fotheringham et al. 2003). The Bandwidth estimated by the model was 366 metres; it took over 2 days for the software to estimate the first model. After the first estimation, the same bandwidth was used in all the other models. The running time is considerably reduced when the bandwidth is provided manually.

6.3.1 Local Model. Value of Square Metre of Construction as Dependent variable

The global model is presented in the table

Table 6-16 Global Model. Dependent variable: Value of square metre of construction

Adjusted r-square		0.309
Parameter	Estimate	t
Intercept	148541.97	66.92
Age	-20.90	-6.91
STOREY	7081.05	20.36
AREA	-19.81	-11.10
Comm12	14485.90	11.12
Crim45	-304.18	-0.44*
AIRPORT	20.16	47.42
ParkRoad	-12.40	-38.76

*Not significant

The global model is presented only to be compared to the local statistics. The R^2 value of the global model is 0.309, and the average R^2 of the 23,344 local models is 0.38; the spatial distributions of the resulting statistics are shown in the maps below.

The distribution of R^2 values are shown in the Map 6-1; the statistics that measure the explanatory power for each of the points. Compared to the goodness of fit (0.309) of the global value, there are large variations in the performance of the model, the local models R^2 varies from 0.22 to 0.98, showing interesting differences in the values across the area of study. There are notorious clusters of areas where the explanatory power of the model is particularly high, and these areas coincide with the most expensive clusters.

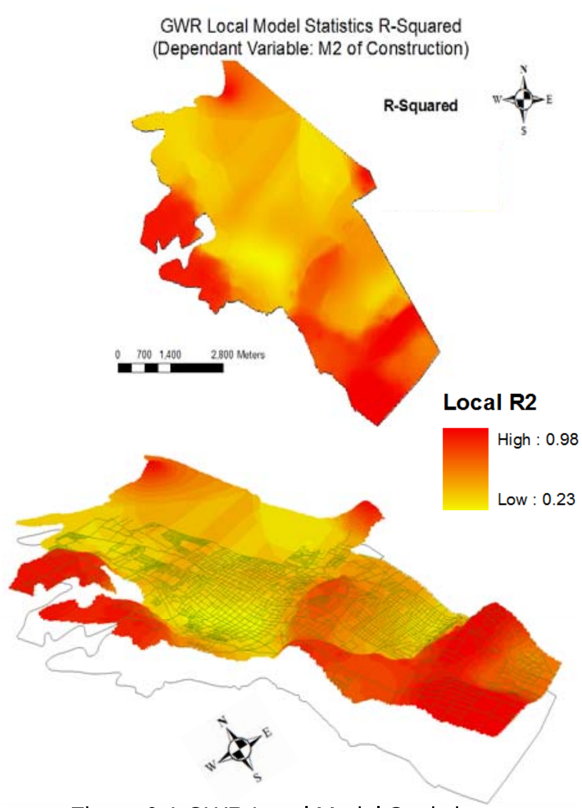

Figure 6-1 GWR Local Model Statistics spatial distribution of local R squared.

The distribution of R^2 values helps to determine the areas that might need more variables to obtain a better replication of the real values. This analysis, led us to conclude that values in less expensive dwellings are more difficult to replicate than the values in expensive ones; values in expensive clusters seem to be well explained by location factors. On the other hand, there is evidently a suggestion that the variables presented in the model fail to account adequately for the value of the properties in poorer areas. This is interesting, and it makes sense in the fact that in Bogota, less expensive properties are less homogeneous; hence, there are more complex factors that have to be taken into account in order to predict their prices.

The second type of statistics to be discussed is the coefficient of the variables. Even though all of the variables are not presented here because of space limitations, it is possible to map the values for each parameter estimate from the GWR model. Only three variables are discussed and reported, and can be seen in the Figure 6-2.

Another important element to notice is that the distribution of values of coefficients in some areas is not even over the space. This is a sign that there is non-stationarity of the coefficients over the space. This can be seen in the Figure 6-2.

For the variable park and main roads, it seems that the properties that are closer to them have the strongest negative impact, meaning that for being close, its price is elevated. That does not seem to be the same in other areas, where being far from the park and main roads seems to increase the price, especially in the high value spots located in the northwest part of the District. This may have a particular explanation, which is that despite being far from the main roads, there are high value clusters which may have some benefit for being far from the main roads. What is clear from the surface result map is that the park road coefficients vary over space, and the assumption of uniformity of the global model does not appear to be appropriate.

The proximity to the airport coefficient has a similar behaviour to the park roads coefficients. The properties that are closer to the airport have a higher positive coefficient, meaning that the farther to the airport the more expensive. In the south east border of the District, where the most expensive properties are, there is a patch of properties with a negative coefficient value, meaning that farther is less expensive. This is neither consistent with the other global models nor consistent with the value tendencies or common sense.

The last coefficients to be discussed are the ones for age of the property. Following the common sense of real estate for the majority of cases, older properties are less expensive. This is true for most of the District, with the exception of a few clustered areas with negative values. This means that in that specific area older properties are more expensive, which might be the result of the first residential developments in the expensive area that possibly gain value with the years as valorization. Another explanation can be that these particular properties not only share the age but also some attractive conditions that make them more expensive.

Again, the spatial variation in the parameters is significant because there are clusters of tendencies. The values of the coefficients for high price properties tend to cluster together, as well as the values of coefficients for low price properties. This element fosters the hypothesis that low price properties and high price properties may perhaps need different variables and the impact of each variable is different.

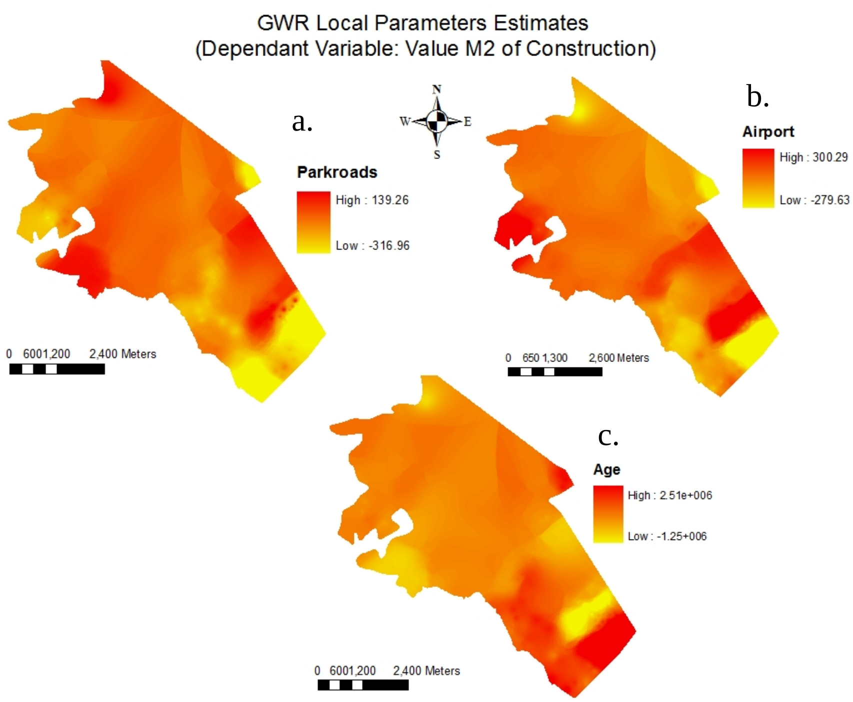

Figure 6-2 GWR Local Parameters Estimates (Dependent Variable: M2 of construction), spatial distribution of local parameters: a. distance to park roads variable, b. proximity to airport, c. age of construction.

The last discussion is regarding the predicted values using GWR and the resulting residuals values Figure 6-3. The predicted surface value has a more even distribution than the real one and the general outcome is similar but with clearer high or low price clusters. The original value map has a more heterogeneous behaviour, with more peaks (see Figure 5-2). These differences can be seen in the residual surface map Figure 6-4). The residual surface map reports the differences between real and estimated values.

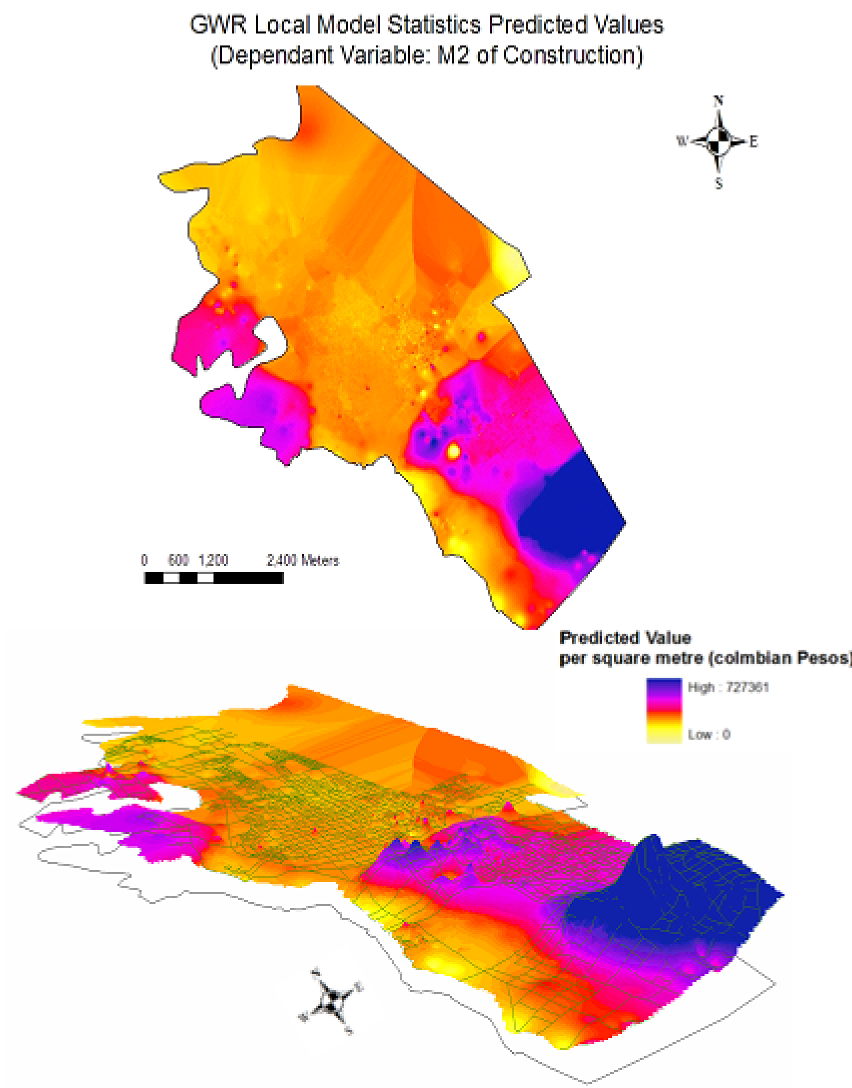

Figure 6-3 Predicted Values of M2 of construction using GWR

The residual values depict the accuracy with which the model replicates the real values, and it shows that the differences at some specific points are large. The residuals are strongly related to the R^2. It is also possible to identify clusters where the real values are not replicated as well. The residuals map also shows variability in the residuals across the space.

It is important to point out that some of the differences replicating the values are due to lack of some important variables that represent attributes that characterize the properties.

6.4 Chapter Conclusions

Global models for the value of square metre of construction have a low R^2 value; the hypothesized reasons for this are:

- Location variables do not account enough for the price of the property, meaning that variables that characterize the property with other attributes are needed in order to increase the explanatory power of the models. Residential price models require a number of details that characterize each property, and these were not available for this study; therefore, the explanatory power of the models was limited.
- Location variables have more explanatory power for the land value models because land value does not need to be characterized by specific property attributes.

GWR offers a complementary approach to property valuation assessment, which explicitly accounts for the local influences and overcomes, to some extent, the shortcoming of global models.

The main conclusion hypothesis from the local statistics models is that it seems that values of more expensive houses are better explained by location variables. Less expensive dwellings require more variables to explain its value. This conclusion is important since it could mean that for more expensive properties the location is more important than for less expensive ones. This may be related to the dynamics of developing cities, where the best locations are taken for expensive developments, and the less expensive houses are relegated to the less strategic locations. Therefore, the less expensive residences all share a poor access to facilities and communication, and subsequently, the value is mainly determined for other variables different than the location variables that were not considered in this thesis.

This conclusion would have never been possible with global statistics models. The resulting surface maps with local statistics show that the coefficients vary over space, and allow one to notice that models for the less expensive areas had lower R^2, and that the location variables had less impact on these areas. The models are less explanatory mainly because most of the variables are proximity attributes that seem to have less effect in less expensive properties. For the more expensive properties the R^2 is higher, which can be interpreted as that proximity variables are highly important for high price clusters.

The local land value models have a better predicting capabilities, this might be because they do not need the detail of the construction characteristics. Despite being spatial heterogeneous, the spatial differences in the parameters is not as notorious as for the construction models.

Despite the advantages, the use of GWR raises some concerns. The local multicollinearity is still valid with local models. The selection of an appropriate method to visualize the results is important, and the awareness that an interpolation procedure might carry its own errors is also important. In practical terms, the GWR are computationally burdensome and require large amounts of processing time.

The visualization capabilities of GIS demonstrated to be helpful to visualize the GWR results. The use of ArcScene for 3D representations helps the presentation and visualization but add some perception errors, since vertical exaggeration is required.

7 CONCLUSIONS AND DISCUSSIONS

Real estate mass valuation models are fundamental for local and national governments, since authorities require the valuation of large number of properties for taxation purposes. A mass valuation appraisal has to be homogeneous, transparent, accurate, fair, and relatively inexpensive. GIS has proved to be a fundamental tool for modern real estate mass valuation, GIS assures objectivity in the process and is a complement to traditional econometric models. In mass appraisal, GIS is valuable as a visualization tool and most importantly to generate spatial variables that in the past were not possible without GIS.

The state of the art instrument for valuation of real estate is the integration of Computer Assisted Mass Appraisal (CAMA) and GIS. In traditional models most of the structural characteristics are relatively easy to measure, but location characteristics, such as distance to major transportation arteries, proximity to shopping, crime rates, and so on are more difficult to analyze. GIS has contributed to the development of spatial variables and proximity attributes.

The Local Statistics approach explored in this thesis can be valuable for real estate property valuation, since traditional mass valuation models are global statistics models that ignore potential variations across space. In contrast, local approaches analyze regional variation in space, and the impact of each variable in the value of real estate seems to be spatially non-stationary.

7.1 Specific conclusions

1. The visual interpretation analysis of the variables with surface value maps offers important possibilities for interpretation, and most importantly, it helps to interpret the models and to identify odd estimates.

2. The global models reported strong correlation among the location variables and the coefficient of determination (R^2) between 0.3 and 0.6, meaning that location

variables do not account enough for the price of constructed properties. The goodness of fit was not high because most of the variables were based on proximity attributes. Nevertheless, the location variables had more explanatory power for the land value models, since land value does not need to be characterized by specific property attributes.

3. In the global models, the location variables that have the higher impact on land and property values are proximity to an airport, proximity to a motel, proximity to park and roads, and shopping, which coincide with the visual interpretation hypothesis. But the interpretation of the values of the resulting coefficients must be cautious, because of the strong correlation between the variables, therefore in future models it is not recommendable to use all of them at the same time.

4. Crime, measured as events (number of burglaries and locations), is a complicated variables to analyze since the locations were spread all over the study area. In this study the impact of high crime density was not clear, neither with visual interpretation nor with the models. This is, by some means, consistent with the literature which argues that crime itself is not price determinant.

5. The residential property values were highly clustered, as indicated by the Moran's I index (Figure 7-1) performed for the value of construction for the 23,344 points.

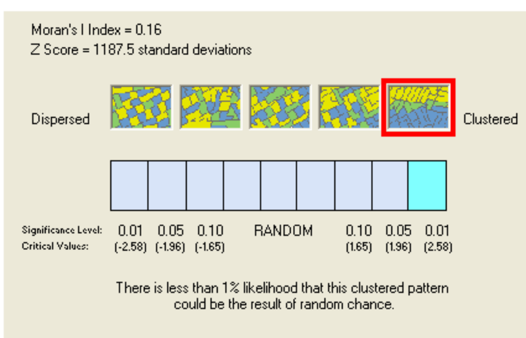

Figure 7-1 Moran's Index for construction values in the area of study

6. The spatial variation in the parameters is significant, because there are clusters of tendencies. High price properties tend to cluster together, as well as the low

price properties. This strong clustered distribution makes it ideal to use local spatial statistics to analyze the different values across the study area. The use of Geographical Weighted Regression (GWR) with GIS as a visual technique revealed some interesting patterns, which leads to an interesting suggestion:

- o Low price properties are more difficult to replicate than high price properties. The local statistics result show that the values of expensive clusters seem to be well explained by location factors. In contrast, they suggest that the variables presented in the model fail to account adequately for the value of the less expensive clusters. Hence, it can also be interpreted that location variables do not have too much impact on low price clusters.

This conclusion may be explained by various reasons:

- The best locations are taken for expensive developments, and the less expensive houses are relegated to the less strategic locations. Therefore, the less expensive residences all share poor access to facilities and communication, and subsequently, the value is mainly determined by other variables different than the location
- In Bogota, less expensive properties are less homogeneous; thus, there are more complex factors that have to take into account in order to predict their prices.

7. In terms of applicability of Local Statistics for real-world, GWR can potentially be used for mass appraisal: in the case of this study, 23,344 properties were assessed and the values of all were predicted. Therefore, the local statistics are worthy to be considered. Nevertheless, compared to global models, few academic experiences have used Local Statistics for real estate, and to the knowledge of the author, there are none in real world applications. The main reason could be that Local Spatial Statistics methods are computationally cumbersome, which make them difficult to integrate into big scale applications: for instance, a linear regression estimation for the data of this study takes less that 5 seconds, and the GWR estimation takes 3 days. In the near future, new methods or more efficient software, will allow the spread of local spatial statistical

models among practitioners. The new version of ESRI ArcGIS, version 9.3, have a GWR module (ESRI 2008).

8. The visualization capabilities of GIS demonstrated to be helpful to construct value maps for visual interpretation hypothesis. In addition, the use of ArcScene for 3D representation helps visualization (e.g. Figures 6-1), but might induce perception errors because of the use of vertical exaggeration. GIS also proved to be fundamental to present the results of local spatial statistical models, since each point has its own model. The best way to report the statistics for each point, local statistics, is with the use of GIS; furthermore, the possibility of using 3D surfaces can be helpful in analyzing the spatial variability of statistical parameters.

7.2 Directions for future research

A number of directions for future research can be identified in this study.

The first, and probably the most important, is to extrapolate the model to a sample taking into account the entire city, and not limit the study only to location variables, but also to include specific property attributes. This study showed that some of the most important location variables are strongly correlated; a careful selection of the most representative and less-correlated location variables can be done.

It would be ideal to compare global with local regression models for the entire city; it is envisioned that at a city scale there are more heterogeneous properties and the spatial differences are more notorious. An extensive study of the entire city might help to foster or discourage the hypothesis that prices in less expensive clusters are more difficult to replicate with models. A careful study will be necessary to identify possible variables that affect the values of the less expensive areas. It would also be ideal to have a database with market values to make a comparison.

8 REFERENCES

Allison, P. (1999). Multiple Regression: A Primer. Thousand Oaks, CA, Pine Forge Press.

Anselin, L. (1990). "Some robust approaches to testing and estimation in spatial econometrics." Regional Science and Urban Economics 20: 141-163.

Anselin, L. (1992). "Spatial statistical analysis and geographic information systems." The annals of Regional Science 26: 19-33.

Anselin, L., J. Cohen, D. Cook, W. Gorr and G. Tita (2000). "Spatial Analyses of Crime." Criminal Justice 4: 213-262.

Basu, S. and T. Thibodeau (1998). "Analysis of Spatial Autocorrelation in House Prices." Journal of Real Estate Finance and Economics 17(1): 61-85.

Bitter, C., G. Mullingan and S. Dall'erba (2007). "Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method." Journal of Geographical Systems 9: 7-27.

Block, C., M. Dabdoub and S. Fregly (1995). Crime Analysis Through Computer Mapping. Washington, D.C, Police Executive Research Forum.

Boba, R. (2005). Crime Analysis and Crime Mapping. Thousand Oaks, California, Sage Publications, Inc.

Bogota, A. M. d. (2008). "Indicadores Financieros." www.Bogota.gov.co Retrieved May 2008.

Bogota Turismo. (2008). "http://www.bogotaturismo.gov.co/atractivos/parques/zona_occidente.php#simon_bolivar." Retrieved June 2008.

Borst, R. (2006). Location Models Revisited: Past, Present and Future. Integrating GIS & CAMA Conference, Orlando, Florida, February 2006.

Bowes, D. and K. Ihlanfeldt (2001). "Identifying the Impacts of Rail Transit Stations on Residential Property Values." Journal of Urban Economics 50: 1-25.

Box, G. E. P. and G. M. Jenkins (1968). "Some advantages in forecasting and control I." Applied Statistics 17(2): 91-109.

Box, G. E. P. and J. F. MacGregor (1974). "Some advantages in forecasting and control II." Applied Statistics 23(2): 158-179.

Branas-Garza, P., J. Rodero and J. Presley (2002). "The North-South Divide and House Price Island: The Case of Cordoba (Spain)." European Journal of Housing Policy 2(1): 45-63.

Bustamante, L. and D. Erba (2008). Catastro Territorial en la Republica de Colombia. El Catastro Territorial en America Latina y el Caribe. D. A. Erba, Lincoln Institute of Land Policy: 127 - 140.

Bustamante, L. and N. Gaviria (2004). "The Bogotá Cadastre: An Example of a Multipurpose Cadastre." Land Lines - Lincoln Institute of Land Policy 16(2).

Catastro. (2005). "http://www.catastrobogota.gov.co/Consulte/Estadisticas.htm, webpage Cadastre Bogota." Retrieved May 2008.

Chambers, D. (1992). "The Racial Housing Price Differential and Racially Transitional Neighborhoods." Journal of Urban Economics 32(2): 214-232.

Chesire, P. and S. Sheppard (1998). "Estimating the Demand for Housing, Land, and Neighbourhood Characteristics." Oxford Bulletin of Economics and Statistics 60(3): 357-382.

Chesire, P. and S. Sheppard (2004). "Introduction to Feature: The Price of Access to Better Neighbourhoods." The Economic Journal 114: 391-396.

Clapp, J. and S. Ross (2004). "Schools and Housing Markets: An examination of school segregation and performance in Conneticut." The Economic Journal 114: 2004.

Crawford, G. W. and M. C. Fratantoni (2003). "Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices." Real Estate Economics 31(2): 223-243.

d'Amato, M. (2002). "Appraising property with rough set theory." Journal of property Investment and Finance 20(4): 406-418.

d'Amato, M. (2004). "A comparison between MRA and Rough Set Theory for mass appraisal. A case in Bari." International Journal of Strategic Property Management 8(4): 205-217.

DANE. (2008). "Censo General." www.DANE.com Retrieved May 2008.

Departamento Administrativo de Catastro (2001). MANUAL CÁLCULO EFECTO PLUSVALÍA. Alcaldia Mayor de Bogota.

Des Rosiers, F., A. Lagana and M. Theriault (2000). "Size and proximity effects of primary schools on surrounding house values." Journal of Property Research 18(2): 149-168.

Des Rosiers, F., A. Lagana, M. Theriault and M. Beaudoin (1996). "Shopping centres and house values: an empirical investigation." Journal of Property Valuation & Investment 14(4): 41-62.

Des Rosiers, F., M. Theriault and P. Villeneuve (2000). "Sorting out access and neighbourhood factors in hedonic price modelling." Journal of Property Investment and Finance 18(3): 291-315.

Dubin, R. (1988). "Estimation of regression coefficients in the presence of spatially autocorrelated error terms." The Review of Economics and Statistics 70(3): 466-474.

Dubin, R. (1992). "Spatial autocorrelation and neighborhood quality." Regional Science and Urban Economics 22: 433-452.

Eckert, J. (1990). Property Appraisal and Assessment Administration. Chicago, IL, International Association of Assessing Officers.

ESRI (2008). What's New in ArcGIS® 9.3. http://webhelp.esri.com/arcgisdesktop/9.3/pdf/Whats_New_In_ArcGIS_93.pdf. Redlands, CA.

ET GeoWizards. (2008). "ET GeoWizards 9.7." http://www.ian-ko.com/.

Federal Land Cadastre Service of Russia (2001). Land (Real Estate) Mass Valuation Systems for Taxation Purposes in Europe. F. L. C. S. o. R.-U. E. W. P. o. L. Administration.

Fik, T., D. Ling and G. Mullingan (2003). "Modeling spatial variation in housing prices: A variable interaction approach." Real Estate Economics 31(4): 623-646.

Fischer, M. and P. Nijkamp (1992). "Geographic information systems and spatial analysis." The annals of Regional Science 26: 3-17.

Fisher, J. and R. Martin (2004). Income Property Valuation. Chicago, IL., Dearborn Financial Plubishing, Inc.

Fotheringham, S. (1997). "Trends in quantitative methods I: stressing the local." Progress in Human Geography 21(1): 88-96.

Fotheringham, S., C. Brunsdon and M. Charlton (2002). Geographically Weighted Regression: the analysis of spatially varying relationships. West Sussex, England, John Wiley & Sons.

Fotheringham, S., M. Charlton and C. Brunsdon (1998). "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis." Environment and Planning A 30: 1905-1927.

Fotheringham, S., M. Charlton and C. Brunsdon (2003). GWR 3 Software for Geographically Weighted Regression: Manual.

German, J., D. Robinson and J. Youngman (2000). "Traditional Methods and New Approaches to Land Valuation." Land Lines - Lincoln Institute of Land Policy 12(4): 4-5.

Gibbons, S. (2004). "The Cost of Urban Property Crime." The Economic Journal 114: 441-463.

Gloudemans, R. and E. Montgomery (2006). Valuation with SPSS and Spatialest in Northern Ireland. Integrating GIS & CAMA Conference, Orlando, Florida, February 2006.

Hernandez, T., M. Yeates and T. Lea (2003). Residential Property Valuation: An Application of Geographically Weighted Regression (GWR). CENTRE FOR THE STUDY OF COMMERCIAL ACTIVITY. Toronto, Ryerson University, Toronto, ON. RL 2003-05: 1-12.

Kryvobokov, M. (2005). "Estimating the weights of location attributes with the Analytic Hierarchy Process in Donestsk, Ukraine." Nordic Journal of Surveying and Real Estate Research 2(2): 5-29.

Kryvobokov, M. (2007). "What location attributes are the most important for market value?: Extraction of attributes from regression models." Property Management 25(3): 257-286.

Kryvobokov, M. and M. Wilhelmsson (2007). "Analysing location attributes with a hedonic model for apartment prices in Donetsk, Ukraine." Journal of Strategic Property Management 11: 157-178.

Lake, I., A. Lovett, I. Bateman and B. Day (2000). "Using GIS and large-scale digital data to implement hedonic pricing studies." International Journal of Geographical Information Science 14(6): 521-541.

Lentz, G. and K. Wang (1998). "Residential appraisal and the lending process: A survey of issues." Journal of Real Estate Research 15: 11-39.

Lynch, A. and D. Rasmussen (2001). "Measuring the impact of crime on house prices." Applied Economics 33: 1981-1989.

Malme, J. (2004). "Mass Valuation for Land Taxation in Transitional Economies." Land Lines - Lincoln Institute of Land Policy 16(2).

McCluskey, W., W. Deddis, A. Mannis, D. McBurney and R. Borst (1997). "Interactive application of computer assisted mass appraisal and geographic information systems." Journal of Property Valuation & Investment 15(5): 448-465.

Milam, J. and F. Mitchell (2003). Integrating a Fully Functional CAMA/GIS System in the Field - The Dallas Central Appraisal District Mars Project. Integrating GIS & CAMA Conference, Columbus, Ohio, March 2003.

Pace, K., C. F. Sirmans and C. Slawson (2001). Automated Valuation Models. AH Dordrecht, The Netherlands, Kluwer

Pagourtzi, E., V. Assimakopoulos, T. JHatzichristos and N. French (2003). "Real estate appraisal: A review of valuation methods." Journal of property Investment and Finance 21(4): 383-401.

Pavlov, A. (2000). "Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets." Real Estate Economics 28(2): 249-283.

Pawlak, Z. and R. Slowinski (1994). "Decision Analysis Using Rough Sets." International Transactions in Operational Research 1(1): 107-114.

Pompe, J. and J. Rinehart (1995). "Beach Quality and the Enhancement of Recreational Property Values." Journal of Leisure Research 27(2): 143-154.

Powe, N., G. Garrod, C. Brunsdon and K. Willis (1997). "Using a geographic information system to estimate an hedonic price model of the benefits of woodland access." Forestry 70(2): 139-149.

Rodriguez, M., C. F. Sirmans and A. P. Marks (1995). ""Using geographic information systems to improve real estate analysis." The Journal of Real Estate 10(2): 163-73.

Secretaria de Gobierno Bogota (2006). Diagnostico de Seguridad Localidad de Fontibon.

Thaler, R. (1978). "A note on the Value of Crime Control: Evidence from the Property Market." Journal of Urban Economics 5: 137-145.

Thompson, M. (2004). "The Bogotá Cadastre: An Assessment " Land Lines - Lincoln Institute of Land Policy 16(2).

Tse, R. (2002). "Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach." Urban Studies 39: 1165-1180.

URISA. (2008). "http://www.urisa.org/gis_cama."   Retrieved May 2008.

Vazquez, J. (2006). Catastro Benchmarking. L. J. Vazquez, Republica de Panama.

Welcome, P. and L. Clark (2003). Automated Valuation Models. Integrating GIS & CAMA Conference, April 2003, Columbus, Ohio.

Whiterell, D. (2007). Automated Valuation Models by Model Building Practitioners. Integrating GIS & CAMA Conference, Las Vegas, Nevada, March 2007.

Wilson, P., J. Okunev, C. Ellis and D. Higgins (2000). "Comparing univariate forecasting techniques in property markets." Journal of Real Estate Portfolio Management 6(3): 283-306.

Wyatt, P. (1996). "Using a geographical information system for property valuation." Journal of Property Valuation & Investment 14(1): 67-79.

Wyatt, P. (1997). "The development of a GIS-based property information system for real estate valuation." International Journal of Geographical Information Science 11(5): 435-450.

Wyatt, P. and M. Ralphs (2002). GIS in Land and Property Management. London, Spon Press - Taylor & Francis Group.

Yang, Z. (2001). "An application of the hedonic price model with uncertain attribute The case of the People's Republic of China." Property Management 19(1): 50-63.

Yavas, A., V. Dokmeci and Z. Onder (2003). "External Factors, Housing Values, and Rents: Evidence from Survey Data." Journal of Housing Research 14(1): 83-101.

Series from Lund University's Geographical Departments

Master Thesis in Geographical Information Science (LUMA-GIS)

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland. (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research. (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing. (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Colombia. (2009).