ANALYSIS OF THE IMPACT OF LOCATION FACTORS DERIVED FROM DIFFERENT TECHNIQUES ON THE PREDICTIVE ACCURACY OF CAMA MODELS.

THEMBA MUSEKIWA

Thesis presented for the degree of Masters of Engineering

In the Department of Architecture, Planning and Geomatics
University of Cape Town

August 2004

DECLARATION

I hereby:

(a) grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research.

(b) declare that:

(i) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from the Research and Development Team, from the Cape Town City Council based in the Department of Property Valuation.

(ii) neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree in the University or any other University.

(iii) I am now presenting the thesis for examination for the Degree of Masters in Engineering.

Themba Musekiwa **August 2004**

DEDICATION

"to all members of the Musekiwa clan, without you I would have never got this far"

ABSTRACT

Author: Themba Musekiwa

Address: 5-157 Mpopoma, Bulawayo, Zimbabwe

Thesis Title: Analysis of the impact of location factors derived from different techniques have on the predictive accuracy of CAMA models.

Date: August 2004

The primary objective of this research is to assess the impact of using different techniques for accounting for locational effects on the estimated market values of properties in the City of Cape Town. Property data from the Cape Town City Council's (CTCC) General Valuation 2000 (GV2000) Project is used for creation of response surfaces and development of computer assisted mass appraisal (CAMA) models. The dataset is split into a modelling sample and a test sample, in the ratio 80:20 respectively. The modelling sample is used for development of three base models and for creation of response surfaces, whereas the test sample is used for conducting validity and reliability checks. Locational effects are examined using four approaches; homogenous neighbourhoods, stratification based on a geographical feature that is Table Mountain, location value response surfaces and delineated neighbourhoods. The above-mentioned techniques have been tried and tested in most jurisdictions except for trend surface analysis (TSA), which has only been used in Brazil. The response surfaces are built using two *zscores*: one derived from the sale price per square metre of the properties and the other from the property characteristics. Location factors estimated from these surfaces are then incorporated into the base models thereby allowing for statistical analysis to be conducted on the results produced by the models. Locational adjustments are found to be very significant in the estimation of market values of properties, with the best technique for generating location factors for the City of Cape Town being ordinary kriging.

ACKNOWLEDGEMENTS

City of Cape Town Valuation Management:

I would like to thank *Mr. Christopher Gavor* and *Mr. Emil Weichardt* for enabling me to work with the CAMA team throughout the research.

CAMA Analysts:

Mr. Ian Smoothey, Mr. Lewellyn Louw and *Ms. Yolanda Crisp* - your input to this research was invaluable.

Fellow Geomatics Postgraduates:

Thank you for the contributions you made during the research meetings.

Ms Jenny Whittal:

Big thanks for taking me on board, as well as for the support and guidance you have provided throughout the research.

Funding:

This research was supported financially and in kind by the following organisations:

The City of Cape Town, Property Valuations Section

The University of Cape Town

The Lincoln Institute of Land Policy

CONTENTS

List of Figures

List of Tables

List of Appendices

CHAPTER ONE

Research Overview

Chapter one provides the reader with an overview of the research. The problem statement, which is the motive for the research, goals to be achieved and the limitations of the research, are briefly tackled in this chapter.

1.1 Introduction

The National Association of Tax Assessing Officers, today known as the International Association of Assessing Officers (IAAO), was formed in 1934 with the intention of improving mass appraisal and promoting consistency with regard to appraisal practices (Gloudemans, 1999). Isakson (2001) states that credit should be given to Court for his 1939 article on the modelling of the price of an automobile based on its features, for this idea could have lead to the concept of mass appraisal. A number of statistical and mass appraisal packages where introduced in the mid 1970's, although some Californian jurisdictions had been using multiple regression analysis (MRA) for about a decade (Eckert, 1990). This resulted in the automation of mass appraisal, and subsequently led to the development of Computer Assisted Mass Appraisal (CAMA) systems (Gloudemans, 1999).

CAMA can be thought of as a systematic process of appraising a specific type of real property using statistical analysis procedures such as MRA (Eckert, 1990) Initial implementation of a CAMA system can be very expensive, but in the long run it results in reduced costs compared to single property appraisal, especially with regard to reappraisal of an entire jurisdiction. Long term cost efficiency, consistent and accurate modelling of property value, and time efficiency are the main pros for the implementation of a CAMA system. It is, however, a new technology and its successful implementation requires high levels of expertise and experience.

The value of real property is a function of many variables such as its condition, neighbourhood and physical extent, but the most influential is location. Knowledge of the effect of location, although critical in determining the value of a property, comes with a price in that it cannot be directly measured in the real estate market like the size of the property. Immobility of the property results in its value being spatially dependent, that is it depends on location attributes (Gonzalez et. al., 2002). The above statement is supported by the fact that nearby and similar properties have similar market values, and increasing the distance between them will reduce this similarity (Dubin, 1992; Dubin and Sung, 1987). The major problem usually experienced in CAMA is the issue of measuring locational effects over an entire jurisdiction (Gonzalez et. al., 2002).

During the market value-modelling phase, a location factor is included as an independent variable, thereby allowing for trends in the sale price to be captured due to location. Major cities are structured in such a manner that they consists of multiple centres of economic influence such as shopping complexes thereby complicating the process of modelling location (McCluskey and Deddis, 2002). In fact there are many factors that may further complicate the modelling of location and for residential property this could be proximity to natural features, open spaces and so forth, as well as distance from undesirable areas such as industrial sites. The traditional approach to assessing locational effects has been through delineated or fixed neighbourhoods, but the problem with this technique is that it does not adequately address the issue of micro-variations within neighbourhoods in a single model and also the property market is constantly changing (Dubin, 1992; Galllimore et al, 1996; Robinson, 1997; Siu and Yu, 2001). Today, location is analysed using complex techniques such as artificial neural networks (ANN) and location value response surfaces (LVRS).

1.2 Problem Statement

The social, political and economic dynamics in the City of Cape Town (CCT) display spatial characteristics, which lead to an extremely complex property market with highly significant spatial variation. In general, this differs from cities in the developed countries where most neighbourhoods have much greater levels of homogeneity. Although nearby properties should generally have similar characteristics and hence value, in reality this seems not to be the case in some parts of Cape Town.

Developing models for estimating market values under such conditions creates problems such as undervaluing of properties and vice versa. This generally leads to market value estimates not being sufficiently representative of the real situation in the property market. It is hoped that by improving the modelling of location, better estimates can be obtained thereby leading to a closer realization of the goals of revaluation and the use of CAMA in the valuation process.

Surveyors have a good knowledge of applied mathematics and statistics, as well as spatial data modelling. This combination of skills and knowledge is likely to contribute to improving the modelling of the effect of location on property value, and hence would assist in the achievement of the broader goals of the City for future valuation cycles.

1.3 Research Objectives

1.3.1 Primary Objective

The primary research objective is to investigate and analyse the use of various techniques including those employing geographic information systems (GIS) for modelling the influence of location on the estimation of the market value derived from Computer Assisted Mass Appraisal (CAMA) models for single residential real property.

1.3.2 Secondary Objectives

- Provide input to the City for future valuations on improving the market value estimates through improving location modelling.
- Capacity building in CAMA and research in fiscal cadastres at the University of Cape Town (UCT) and within South Africa.

1.4 Hypothesis

1.4.1 Primary Hypothesis

Different techniques can be used for modelling the effect of location on market value estimates, and some techniques produce better results than others.

1.4.2 Secondary Hypothesis

- Identifying and using a good technique of modelling the effect of location in CAMA modelling will improve the estimation of market values. This is indicated by improvements in the sales ratio i.e. minimising the variations between the market value estimates and the sales prices.

1.5 Research Questions

- What is the influence of location on the modelling of market values of residential real property using CAMA and GIS?
- What effect do different location analysis techniques have on the estimated market values?
- Does a spatial trend exist with regard to the sale price and in which direction is it oriented?
- Based on the answers to the above, what recommendations can be made to the CTCC for future valuations regarding the best technique of location modelling?

1.6 Scope of the Research

1.6.1 Data

The sales and other property data used for this research are restricted to the data used for the 2000 General Valuation (GV2000) Project of the City of Cape Town. The size and distribution of the dataset will be investigated thoroughly so as to ascertain whether it is adequate to enable conclusive results to be reached. The sales data obtained will play a major role in determining whether the models produced for estimating market values are of high quality. This is due to the fact that we are dealing with sale transactions that occur between people, and most of the time the price paid for a property is subjective. Another reason why sales data plays a major role is that if a neighbourhood has no sales data then it means that particular neighbourhood will not play a role in the development of locational effects for that area. This implies that the estimates produced for that particular neighbourhood would be constrained to regional trends and effects, and not local variations. The processes used in the GV2000 Project for data cleaning and linearisation will be adopted as is in this research, since the main focus of the research is to model locational effects. Verification of sales assumed to be outliers and estimated market values in the field will not be carried out because the property records being used in this research are for the period 1999 to 2000. Since the research is being conducted for the period 2003 to 2004, it is pointless to carry out these verifications, because within a four-year period changes could have occurred to the properties such as renovations or resale's.

1.6.2 CAMA Models

The CAMA models developed for GV2000 Project will not be used in this research so as to provide the researcher with the opportunity to develop skills in model building. The new models should conform to IAAO standards, but in research such as this where new techniques and combinations are to be tested, this outcome cannot be guaranteed.

1.6.3 Modelling Location

Modelling of location will be examined using four approaches; delineated neighbourhoods, response surfaces, homogenous neighbourhoods and geographic stratification. The main reason for the use of the above approaches is that they deal explicitly with the issue of location and will enable a conclusive result to be obtained on the effect of location on the estimated market values. Another technique exists known as Artificial Neural Networks (ANN), but will not be covered in this research because it is a form of a "black box" that is it does not allow for testing of theory, since an algorithm controls the modelling process and not the researcher. The algorithm selects variables it "thinks" will improve the estimated market values even when the variables cannot be used together. Another reason for its exclusion from the research is that it does not explicitly allow for the analysis of locational effects.

1.6.4 Software

There are two main statistical packages used for statistical analysis and model calibration in CAMA, Statistical Package for the Social Sciences (SPSS) and Number Cruncher Statistical Systems (NCSS). Statistical packages can safely be assumed to deliver similar outputs, as all statistical functions running in the background are identical. For example, a sample mean of a dataset will always be the same regardless of whether a calculator or statistical package is used to compute it. NCSS will be used in this research and its use is justified for a number of reasons:

- The NCSS regression module is simpler to use and less prone to error since variables are specified through a template, whereas SPSS requires that an equation be typed in.
- The owners have made the software readily available at a reasonable cost, whereas the cost of purchasing a single user license of SPSS would have cost US$3000.
- It has a unique feature for generating response surfaces, which is not available in SPSS.

In addition to NCSS, the Microsoft Excel spreadsheet will be used for editing the data as it is more user friendly than NCSS spreadsheet. ArcGIS, an ESRI product, will be used for visualization and response surface analysis because it is readily available to the research team. ArcGIS has a module known as the geostatistical tool, which allows for five different interpolation techniques to be used in building response surfaces. This will be an advantage with respect to the research, as four of these techniques will be used for location modelling. TableCurve 3D software can also be used for response surface modelling but will not be used because it is not available to the research team, and will require more time before the researcher is competent enough to apply it in the project. A comparison cannot be made between TableCurve 3D and ArcGIS, as not much literature or use of the former has been documented except for a demonstration by Ward (2002) where he found that ArcView produced better results to NCSS and TableCurve 3D, because of greater spatial analytical capabilities.

1.7 Research Approach

1.7.1 Literature Review

The best techniques for surface modelling can only be chosen provided one has in-depth knowledge of the pros and cons of the range of different techniques. The value of a property is basically determined by the principles of supply and demand, since valuation principles and practices have their roots in market economics. CAMA and statistical theory are the bases of modern property valuation, and thorough knowledge of these fields is a requirement for useful research in this field. These elements are therefore key to the literature review, which will include a theoretical component to provide the background knowledge for the research, as well as a review of previous research in this field, in order to inform the project.

1.7.2 Data

The data that was used for the General Valuation Project 2000 (GV2000) of the CTCC will be used in this research. It comprises property characteristics for each property such as date of sale. sale price and size of property. After data edits such as cleaning and linearisation of qualitative variables have been performed, the sales file will then be split into two entities. These are the modelling file and the test file in the ratio 80: 20 depending on the number of records available. The modelling file will be used to derive time adjustment parameter; scalar values for qualitative variables and location factors which, in turn, will be applied to the test file. The base models will be derived using the modelling file, with the test-file being used for validation purposes.

1.7.3 Software and Hardware

A number of software packages will be used in the research and these are:

- *Number Cruncher Statistical Systems (NCSS)*- modeling and statistical analysis.
- *Excel and Access* – data editing and storage.
- *ArcGIS v8* – spatial analysis and visualization.

A Pentium 4 (Intel) computer will be used with a processing speed of 2GHz and 512Mb RAM.

1.7.4 Consultation

- *Supervisor* – for guidance through the whole research.
- *CAMA Analysts* – CTCC team providing the technical support.

1.7.5 Modelling

Three base models will be developed in this research and they are:

- *Cape Town* – consists of neighbourhoods within Cape Town MLC
- *Around Table Mountain* – consists of neighbourhoods around Table Mountain
- *Not Around Table Mountain* – consists of neighbourhoods not around Table Mountain

The base models will be used for incorporating location factors derived from the four approaches stated below:

- *Traditional approach* – delineated neighbourhoods
- *Geographic stratification* – based on a physical feature that is Table Mountain
- *Stratification based on mean sale price*– six homogenous neighbourhoods are created using mean sale price intervals.
- *Single model with location adjustments* – involves creation of response surfaces

Traditional Approach

The traditional approach refers to the delineated neighbourhood method, where each neighbourhood is given a single location factor depending on its characteristics. This factor is then applied to all the properties in that neighbourhood and does not take into consideration the variations that may exist between properties within a neighbourhood such as proximity to a school or shopping complex.

Geographic Stratification

In this approach, location is analysed by creating two areas based on a geographic feature known as Table Mountain. The properties surrounding Table Mountain are assigned to a group called around Table Mountain (ATM). The rest of the properties comprising those from the Cape Flats area and the southern suburbs are assigned to a group called not around Table Mountain (NATM).

Stratification based on the Mean Sale Price

The mean sale prices are classified into six intervals thereby giving rise to six homogenous groups. Each neighbourhood with a mean sales price falling within a certain interval is assigned to the corresponding group. Models are then developed for each group so as to enable an assessment to be made about the effect homogeneity has on the estimated market values.

Single Model with Location Adjustments

In this approach location factors generated by the response surfaces will be incorporated into the CAMA models developed in this research. Six response surfaces will be created using the techniques stated below:

Ordinary Kriging

Ordinary kriging is a surface interpolation technique that is generally based on knowledge developed in the geostatistical field using surrounding points (Matheron, 1963; Miller and Khan, 1962). Basic assumptions are that the value surface should be spatially continuous and that value should be a spatially dependent attribute.

Global Polynomial Interpolator

The global polynomial interpolator uses the entire dataset in determining the estimated market value of the unknown property.

Local Polynomial Interpolator

In this method multiple polynomials are used to derive the value of the unsold property and in doing so, this enables the local variations in the sale price to be captured.

Inverse Distance Weight (IDW) Interpolator

The technique estimates the value of the property through a weighting function, assigning greater weights to nearby properties compared to those that are further away from the point of estimation.

Trend Surface Analysis (TSA)

Trend surface analysis is another technique that can be used for location modelling. TSA models consist of polynomial regression models, adding power series of the coordinates to the common property variables (Gonzalez, Soibelman and Formoso, 2002).

Value Influence Centres (VIC)

In this method a contour map of the dependent variable, in our case the sales prices, is created using geographic coordinates, x and y as the position variables. Areas of very high and low value are identified and their locations are noted. The distance from each VIC to each property is computed, and thereby enabling the creation of distance variables. These are then used to compute the location factor corresponding to each property, which will be incorporated into the base model, for further recalibration.

1.8 Structure of Thesis

Chapter One

This gives a brief introduction to the research that is to be carried out and informs the reader of the research questions, problem statement, research hypotheses, objectives and methodology.

Chapter Two

Chapter two consists of the literature review and it will enable a comparison of the different research that has been carried out in the field of computer assisted mass appraisal modelling. The aim of this chapter is to provide a synopsis of the recent research in the field.

Chapter Three

Theory on data analysis, software and modelling will be thoroughly investigated in this chapter. Chapter Three will act as a form of research guideline, as most work that will be carried out in this project will be based on knowledge of the underlying theory.

Chapter Four

This chapter forms the core of the research, as the data exploration and modelling of location will be presented here.

Chapter Five

Analysis of the different techniques used in Chapter Four, and comparison of the results produced, will be the core of this chapter.

Chapter Six

This will be the closing chapter with conclusions and recommendations being presented.

CHAPTER TWO

Review of Previous Research

It is hoped this chapter will enlighten the reader on the work that has been done and the research that continues to be undertaken with regard to techniques used for developing models and generating location factors. A detailed description of the processes and techniques is not given in this section but can be found in Chapter Three, which investigates the theory behind computer assisted mass appraisal (CAMA) and location analysis.

2.1 Modelling

2.1.1 Data

The functional form of a model depends on the modeller, whereas the application and use of the modelling technique relies on the availability of data, its characteristics and goals of the valuation (McCluskey et. al., 1997). A mass appraisal can only be effective and efficient if it is developed from data that is consistent and accurate (McCluskey et. al., 1997; Ward, 2001). Unfortunately, the tendency with most jurisdictions is that they have records that are not consistent; there will be some with missing data, for example (Ward, 2001). Kochetkov (1998) states that the use of non-arms length sales should not be a real problem, however, provided one is working with a huge dataset that emanates from a well-developed real estate market.

2.1.2 Aspect of Location

Spatially, no two properties can be identical, but property valuers will tend to agree on the fact that location is the greatest variable affecting the value of a property (McCluskey, Deddis and Lamont, 2002). Dubin and Sung (1987) demonstrated the existence of economic sub centres such as malls and their effect on the value of real property, thereby assisting in the reinforcement of the multi-centre theory concept by Harris and Ullman (1945). Location can be incorporated into mass appraisal models in a number of ways, one of them being through the delineation of neighbourhoods, or creation of sub-markets. Spatial analysis results in a three dimensional (3D) response surface and provides an alternative to the fixed neighbourhood approach (French et. al., 2003).

Location in Regina, USA was incorporated in the valuation process through delineated neighbourhoods (Figueroa, 1999). Figueroa (1999) carried out research prior to the 2001 reassessment with the hope of using location blind model residuals as a way of measuring location in Regina instead of neighbourhoods. He managed to produce two models each with a coefficient of multiple determination (R^2) of 77.5% and 80.2%, but the city continued using its neighbourhood-based approach. Response surfaces were also used to account for location effects in the pilot study of Cape Town during the period of 1998 to 1999 (Ward, 2001) and were later on adopted in the Cape Town City Council's General Valuation 2000 (GV2000) project. In the research carried out in the City of Calgary, Canada location value response surfaces were shown to be superior to delineated neighbourhoods because of lower coefficients of variation (CoV) and coefficients of dispersion (CoD) (O'Connor, 2002). Recently, Ward *et al* (2002) generated a response surface based on property characteristics instead of the sale price to investigate location effects on the estimated market values using data from Lucas County, Ohio. He found that they produced good results that satisfied the International Association of Assessing Officers (IAAO) set standards as they produced an R^2, CoD and price related differential (PRD) of 94%, 15% and 103% respectively.

2.1.3 Modelling Techniques

Some of the traditional modelling techniques that have been used in the past, and continue to be in use today, are multiple regression analysis (MRA), comparable sales approach (CSA) and adaptive estimation procedure (AEP). Nowadays, advanced methods are being applied because of technological developments. They comprise artificial neural networks (ANN), expert systems, spatial analysis, fuzzy logic and autoregressive integrated moving average (ARIMA). Techniques that have been applied in different countries are as shown in table 2.1, which has been extended from that developed by McCluskey *et al* (1997).

TABLE 2.1 Modelling techniques used in different countries

Country	Department	Modelling technique
Australia (Queensland)	Department of Lands	MRA
Sweden	National Tax Board	MRA
Northern Ireland	Valuer General's Office	MRA, CSA, ANN
Tasmania	Valuation and Lands Agency	MRA, AEP
New Zealand	Valuation New Zealand	MRA
Singapore	Singapore Valuation Authority	Indexation
Hong Kong	Rating and Valuation Dept	MRA, Indexation
Malaysia	Valuation Division	EXPERT
USA	Assessment Offices	MRA, AEP, CSA
British Columbia	BC Assessment Authority	MRA
Brazil	Case Study of Porto Alegre	ANN, TSA, Spatial Analysis
South Africa	City of Cape Town Valuations	Spatial Analysis
USA	New York City Council	Spatial Analysis
USA	Lucas County	Spatial Analysis

2.1.3.1 Advanced Modelling Techniques

Artificial Neural Networks (ANN)

The technique of Artificial Neural Networks (ANN) has been applied to the valuation of residential properties for taxation purposes. McCluskey (1996) applied an ANN to a group of 416 residential properties and managed to obtain a mean absolute error of 7.75%. McCluskey *et al* (1997), with reference to Borst (1995), report that they produced an absolute error of 8.7% after this technique was applied to residential properties. They concluded that the technique of ANN deserves consideration when it comes to mass appraisal. McCluskey *et al* (1997) also states that Tay and Ho (1994) found that ANN generated a mean percentage error of 0.2% and could be used in the valuation processes of gathering and adjusting comparables. Gonzalez *et al* (2002) managed to produce a CoD of 6.7% from the research they conducted in Brazil on the valuation of residential apartments. Research has been conducted to determine the applicability of ANN in mortgage valuations Scott (1988), and in the determination of rental values for commercial property. McCluskey *et al* (1997), with reference to Lenk *et al* (1997) and Worzala *et al* (1995), highlights the fact that despite ANN producing good results, they have limitations in terms of explainability and stability.

Spatial Analysis Methods

Gonzalez *et al* (2002) states that O'Connor and Eichenbaum (1988) used a response surface to derive locational effects for the City of New York. Siu and Yu (2001) applied the technique in Hong Kong and managed to obtain a 15% improvement in the value of R^2 for the model with locational effects. La Rose (1988) produced a global response surface using the triangular irregular network (TIN) method to demonstrate that a global approach produced better results than one created for sub markets.

Trend Surface Analysis (TSA) is a new technique of incorporating location in CAMA models and was demonstrated by Gonzalez *et al* (2002) in Porto Alegre, a town located in the southern part of Brazil. They managed to produce a CoD of 11.2% and an R^2 of 84.5% for their overall model. Gonzalez *et al* (2002) claim that they are the first research team to apply this technique in the field of CAMA modelling, although it has been used for other purposes. Gonzalez *et al* (2002) state that Parker (1981) used it to examine residential land use and household expenditure patterns, while Hembd and Infanger applied it (1981) in the development of models that would predict land values at the urban-rural fringe.

2.1.3.2 Traditional Techniques

Multiple Regression Analysis (MRA)

MRA is the most popular appraisal technique because of its usefulness, simplicity and international recognition as it has been in use for the past four decades. Its popularity continues to grow due to the availability of computer software that performs the necessary calculations. For example different regression modules are inbuilt into standard spreadsheets, thereby allowing for the method to be accessible to non-specialist users (Isakson, 2001). Dubin, Pace and Thibodeau (1999) claim that a number of people have demonstrated the capabilities of MRA in the estimation of market values. Some of its earlier users were Eisenlaner (Eisenlaner, 1968) and Blettner (Blettner, 1969). The preferred type of MRA model is the hybrid model because it is easier to explain to appraisers and ratepayers due to its simple specification, since it takes into account appraisal principles and components in making up the final market value (O'Connor, 2002). The major pitfalls of MRA models can be found in two distinct areas: model specification and robustness of the regression results (Isakson, 2001).

Comparable Sales Approach

Gonzalez *et al* (2002) states that the sales comparison approach is the most used valuation method when dealing with valuation models, which are calibrated using MRA. This approach is appropriate when the real estate market is very active and there are many properties with the same features as the property whose market value is being estimated.

2.2 Software

A number of software packages are utilised in the valuation process, especially spreadsheets and those with statistical analysis capabilities. Lotus, which is a spreadsheet package, was used to carryout MRA calculations in the appraisal of a manufacturing plant (Isakson, 2001). Another spreadsheet package, which is extensively used for data exploration is Microsoft Excel, and was employed in the Cape Town City Council's GV2000 project. In the past, finding software to estimate spatial models was a cumbersome task because of the inaccessible technology, but today due to the advances in software development, there are packages such as Number Cruncher Statistical Systems (NCSS) and Statistical Package for the Social Sciences (SPSS), which can be used directly for model specification and calibration (Dubin; Pace and Thibodeau, 1999). Gonzalez *et al* (2002) used SPSS for regression modelling and Stuttgart Neural Network Simulator (SNNS) for ANN in their research conducted in Porto Alegre. SPSS was used in the research carried out to compare three regression models, that is additive, multiplicative and hybrid, for the City of Calgary's Business Unit, while ArcView's Spatial Analyst was employed for response surface construction (O'Connor, 2002).

There are a number of different software packages, which can be used for exploratory data analysis, but for generating location factors only the three mentioned below are useful (Ward, Guilford, Jones, Pratt and German, 2002);

- TableCurve 3D
- ArcView
- NCSS

Although NCSS is mainly used for MRA model calibration, it has also been employed for almost a decade to generate response surfaces for location factors. Ward *et al* (2002) state that it is the only statistical package with such capabilities, whereas ArcView has been in use for about five years. Ward *et al* (2002) also claim that ArcView is the most precise when it comes to fitting of the sales data followed by TableCurve 3D, the least precise being NCSS. TableCurve 3D was used in developing location factors for Lucas County by the Lucas County Research and Development Staff (Ward et. al., 2002).

CHAPTER THREE

CAMA and Location Analysis Theory

This chapter presents the theory behind CAMA and location analysis. It gradually introduces the theoretical background to mass appraisal and the different approaches used in the valuation of properties. Modelling concepts are then tackled as well as the different types of models that can be used in mass appraisals. The statistical tools used for assessing the effectiveness of the models are also explained in this chapter. The different techniques for generating location factors are discussed with the final section of this chapter focusing on the software packages to be used in this research.

3.1 Mass Appraisal

Mass appraisal is the appraisal of a number of properties, as of a given date, using known procedures and statistical analysis tools (Eckert, 1990). Mass appraisal attempts to replicate market fundamentals (i.e. forces of supply and demand) in the model so that they can be reflected in the estimated market value of the property. Real estate agents, appraisers and public organisations carry out property valuations for a number of reasons such as for property taxation or buying and selling of property. Real estate, in the context of this research, refers to the land and any improvements that are permanently attached to it. Most agencies carrying out property valuations do so with one goal in mind, to produce accurate and reliable market value estimates. The market value of a property can be thought of as the sale price reached upon by a buyer and seller in an open market, provided that they both have sufficient knowledge of the activities of the property market and the transaction occurred under no duress (Eckert, 1990). The sale price of a property is the actual exchange price. A sale price that is not equivalent to the market value may indicate a transaction which is non-arms length, or not representative of the expected sale price of such a property (a realistic price was not paid for the property).

Single appraisal and mass appraisal differ in that the latter is concerned with a group of properties, whereas the former deals with a single property. Market analysis and quality control are dealt with differently in both approaches, but similar appraisal processes are followed. The quality of the models produced using mass appraisal techniques are monitored using statistical tools such as the coefficient of dispersion (CoD) and coefficient of variation (CoV), whereas in single appraisal a property is compared to similar properties that have been sold (i.e. comparable sales). Since mass appraisal targets much larger groups of properties it requires a team of people with multiple skills to work on the process, whereas for single property appraisal only a single person is required. Mass appraisal is preferred to single appraisal because it provides a platform for a uniform and consistent approach in the valuation process, and the potential for cost reduction is high (McCluskey et. al., 1997; Ward, 2001).

The use of computers to manage mass appraisal systems gives rise to Computer Assisted Mass Appraisal (CAMA) systems (Eckert, 1990). CAMA is a system of estimating the value of certain real property using statistical analysis techniques such as Multiple Regression Analysis (MRA) and Adaptive Estimation Procedure (AEP) (Eckert, 1990). Mathematical models, which can be thought of as a function relating the sale price to the property variables, are the "backbone" of CAMA systems. CAMA is a relational database management system (RDBMS), which comprises of four subsystems (Eckert, 1990):

- ***Data Management System*** – this section is responsible for data collection, manipulation and storage.
- ***Valuation System*** – here model specification and calibration are performed.
- ***Performance Analysis System*** – this section monitors the effectiveness of the mass appraisal system through sales ratio analyses.
- ***Administrative System*** – this system is responsible for planning, taxation and appeals from the public.

Mass appraisal comprises of three functions and they are (Eckert, 1990)

- ***Reappraisal*** – it enables an existing appraisal system to be upgraded based on past performance (i.e. sales ratio studies). The ratios assist decision makers as to whether a reappraisal is necessary or not.
- ***Data Maintenance*** – involves capturing and maintaining property data. Re-inspection of properties is conducted after a period of time to check if data in the database is consistent with what is on the ground.
- ***Value Updates*** – it involves the revaluation of property values and is usually conducted on an annual basis, using market trends such as inflation or deflation.

3.2 Data

Data refers to codes, words, numbers or symbols that have been observed or measured in a sequential manner (Gloudemans, 1999). For reliable estimates to be arrived at there is a need to evaluate the quality and quantity of the data used in the estimation process. Data quality is measured by ascertaining how accurate the recorded information, used to estimate values, reflects the current status of the property. Data quantity, however, refers to the amount of data available or that needs to be collected (Gloudemans, 1999). The goal of every jurisdiction with respect to data should be to capture and confirm all the data collected. The tendency nowadays is to rely on the sophistication of technology to conduct property valuations, but modelling with poor data will still yield poor results.

3.2.1 Data Types

Property data can be divided into two distinct groups that is qualitative and quantitative variables.

3.2.1.1 Qualitative Variables

Qualitative variables are nonnumeric that is they tend to be discrete. Most qualitative variables are subjective in that they require some form of judgement to be made. They can be classified into two groups, nominal data and ordinal data.

- Nominal data – it has no order or rank, for example if dealing with the variable, walls of a dwelling, it can have three classes: brick, plastered and wood.
- Ordinal data – the order is important in this group, for example if dealing with the variable, quality of the dwelling, five classes can be identified: excellent, above average, average, below average and poor.

Binary variables are also qualitative, as they are only two forms of response; either a yes or no and they can either be ordinal or nominal depending on the situation.

3.2.1.2 Quantitative Variables

Quantitative variables occur when the observation is in the form of a count or measurement (mathematical, physical or spatial) thereby making them objective.

They can be classified into two groups, discrete data and continuous data.

- Discrete data – generally involves some form of integer count for example number of rooms.
- Continuous data – any value can be recorded provided it falls within a certain interval for example extent of the property.

3.2.2 Exploratory Data Analysis

Exploratory data analysis is conducted so as to enable the modeller to get a "feel" for the data. In order to use only accurate data for modelling, erroneous data needs to be identified and removed. Most errors occur during data collection and capturing, resulting in incorrect property characteristics. This is of particular concern when dealing with qualitative variables relating to property characteristics, as they can be easily incorrectly coded since they are the most subjective. Sales that are non-arms length are the normal type of sale price errors that occur in the data. These sales are identified during the exploratory analysis as well as data that need to be adjusted for personal property (such as furniture) and financing (e.g. the Reconstruction and Development Programme (RDP) properties) (Eckert, 1990). The purpose being to isolate the price paid for real property on the date of analysis, often the statutory assessment date. The data is further analyzed using statistical measures such as central tendency *(median and mean)* and dispersion *(CoD, CoV, standard deviation and variance)* with regard to quantitative data. Frequency distributions and histograms are used to assess the validity of qualitative data. When analyzing two variables simultaneously cross tabulations or scatter diagrams can be used, they enable one to assess whether the variables are correlated.

Correlation (covariation) analysis enables the degree of linearity or interdependency between two variables to be computed and the correlation matrix shows the correlation coefficient between each pair of variables in a dataset. Two forms of correlation exist; positive correlation and negative correlation.

- Positive correlation – implies that two variables increase together for example property prices will increase as the size of the dwelling increases, provided all variables are kept constant such as condition or quality of the dwelling.
- Negative correlation – arises when two variables vary inversely for example the price of a property would be expected to go down as the age of the dwelling increases.

The coefficient of correlation varies from 1 (maximum positive correlation) to −1 (perfect negative correlation). A coefficient of zero implies total statistical independence between the variables, whereas a value close to one signifies dependence. Figure 3.1 illustrates the three forms of correlation that exist. Care has to be taken when interpreting the coefficient of correlation as most a time it can be misleading, for example say a value of 0.45 is obtained, the major issue would be how to interpret it (Rummel, 1976). The best approach to the situation above would be to square the coefficient of correlation, thereby resulting in the squared correlation *(r^2)*, which explains the amount of variance in common between the two variables. Assuming we have three coefficients of 0.9, 0.5 and 0.2, their corresponding coefficients of determination expressed as percentages are 81%, 25% and 4% respectively. The figures above show how misleading the coefficient of correlation values can be when interpreted directly as they imply the existence of more covariation than there is and the situation deteriorates as the correlation values approach zero (Rummel, 1976).

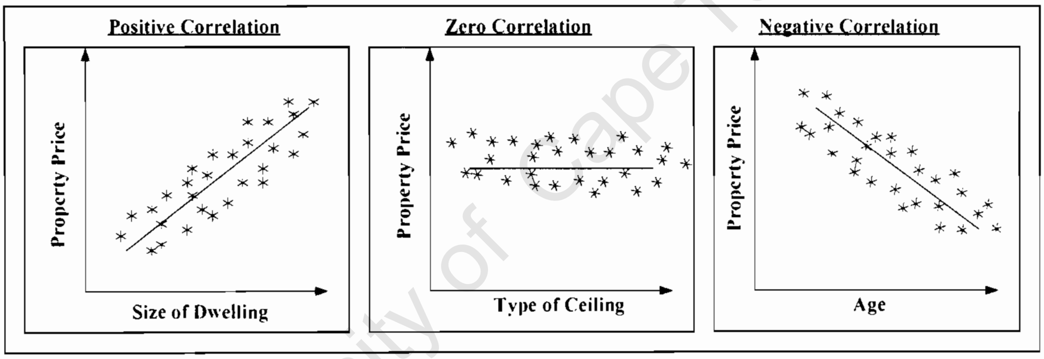


FIGURE 3.1 Types of correlation

The coefficient of correlation between two variables (x_i and x_j) can be derived as shown below (Rummel, 1976).

If each variable has n records, then the mean value of each variable can be computed using equation 3.1.

$$\bar{x}_i = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } \bar{x}_j = \frac{\sum_{j=1}^{n} x_j}{n} \qquad 3.1$$

The variance of each variable is computed as shown in equation 3.2

$$\sigma_i^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2}{n} \text{ and } \sigma_j^2 = \frac{\sum_{j=1}^{n} (x_j - \bar{x}_j)^2}{n} \qquad 3.2$$

The two variables are standardized using equation 3.3 so that their variation is comparable in similar units. This process of standardizing is known as standardization and the transformed values are called standard scores.

$$z_i = \frac{x_i - \bar{x}_i}{\sigma_i} \text{ and } z_j = \frac{x_j - \bar{x}_j}{\sigma_j} \qquad 3.3$$

The product moment correlation coefficient (r_{ij}) between the two variables is computed as shown below and is equal to equation 3.4.

$$r_{ij} = \frac{\sum_{i,j=1}^{n} z_i z_j}{n}$$

$$= \frac{\sum_{i,j=1}^{n} \left(\frac{(x_i - \bar{x}_i)}{\sigma_i} \times \frac{(x_j - \bar{x}_j)}{\sigma_j} \right)}{n}$$

$$= \frac{\sum_{i,j=1}^{n} \left((x_i - \bar{x}_i)(x_j - \bar{x}_j) \right)}{\sigma_i \sigma_j n} \qquad \text{Substituting for } \sigma_i \text{ and } \sigma_j$$

$$= \frac{\sum_{i,j=1}^{n} \left((x_i - \bar{x}_i)(x_j - \bar{x}_j) \right)}{\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2}{n}} \times \sqrt{\frac{\sum_{j=1}^{n} (x_j - \bar{x}_j)^2}{n}} \times n}$$

Therefore the correlation coefficient is;

$$r_{ij} = \frac{\sum_{i,j=1}^{n} \left((x_i - \bar{x}_i)(x_j - \bar{x}_j) \right)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2 \sum_{j=1}^{n} (x_j - \bar{x}_j)^2}} \qquad 3.4$$

3.2.3 Data Transformations

3.2.3.1 Qualitative Data

Quantitative data can be used directly in modelling, but qualitative data must be transformed from its discrete form (e.g. from excellent to poor) into scalar (linear) values for effective manipulation with the quantitative variables, as shown in table 3.1.

TABLE 3.1 Values corresponding to quality of the property

Qualitative		Sale Price		Price per Square Metre		Scalar
Value	Count	Mean	Median	Mean	Median	Value
Excellent	2	2000000	2000000	19643	19643	2.19
Above Average	19	3406451	655000	29806	6384	1.25
Average	250	928100	352500	9840	4524	1
Below Average	31	782303	140000	45277	3125	0.77
Poor	70	185757	13760	3039	237	0.22

The scalar values are derived through market analysis by using the mean or median sale price (SP) and sale price per square meter (SPSM) for each discrete value (class) corresponding to the qualitative variable. The average class is transformed in such a manner that it has a scalar value of zero in additive models and one in multiplicative models. Using the data from table 3.1 and equation 3.5, the transformation process is as shown below.

$$\text{Scalar Value} = \sqrt{\left(\left(\left(\frac{\text{Median SP of each class}}{\text{Median SP of Average Class}}\right)+\left(\frac{\text{Median SPSM of each class}}{\text{Median SPSM of Average Class}}\right)\times 2\right)/3\right)} \qquad 3.5$$

The average class corresponds to the average quality whose median SP is equal to R352 500 and median SPSM is equal to R/m^2 4524. Substituting the median SP and SPSM into equation 3.5;

$$\text{Average Scalar Value} = \sqrt{\frac{\left(\left(\frac{352500}{352500}\right)+\left(\frac{4524}{4524}\right)\times 2\right)}{3}}$$

$$= \sqrt{\frac{((1)+(1)\times 2)}{3}}$$

Therefore the average scalar value is equal to 1.

Transformation of the class equal to excellent with a median SP of R2000000 and median SPSM of R/m^2 19643, follows the same procedure as conducted on the average class and is shown below;

$$\text{Excellent Scalar Value} = \sqrt{\frac{\left(\left(\frac{2000000}{352500}\right)+\left(\frac{19643}{4524}\right)\times 2\right)}{3}}$$

$$= \sqrt{\frac{((5.674)+(4.342)\times 2)}{3}}$$

Therefore the scalar value corresponding to the excellent class is equal to 2.19.

The other classes can be transformed using the same procedure applied in the two examples above. The technique of scaling can be applied in any environment outside the field of CAMA that involves qualitative variables especially if statistical analysis is to be conducted in conjunction with quantitative variables. The transformation of qualitative data into a linear scale also results in the reduction of the number of variables to be used in the modelling process, as seen in the example above instead of dealing with five classes, only one scale (scalar values) is used. The advantage of using this technique arises especially when dealing with a large number of qualitative variables as it minimises data storage and processing times. Although it is a good method of transforming the data, caution must be exercised when using these values because the mean sale prices may reflect other influences especially if outliers have more influence in its derivation (Eckert, 1990).

3.2.3.2 Quantitative Data

Transformations conducted on quantitative data help improve the usefulness, for example if dealing with geographical coordinates of great magnitudes, they may need to be scaled down through division or subtraction of some constant value so as to enhance processing times. The common mathematical transformations used are;

Reciprocal Transformation

The integer one is divided by a variable's numerical value. A typical scenario where this transformation would be used is in deriving weights based on the inverse distance weighting function (see 3.8.1.1a) to estimate the market value of an unsold property based on those that have sold.

Exponential Transformation

A variable is raised to a power or exponent so as to increase or reduce its original value. For example if dealing with property prices that are too high, a power less than the value of one (i.e. a fraction) can be applied so as to reduce their values to manageable levels.

Logarithmic Transformation

The logarithmic transformation is mainly used to transform a variable with a skewed distribution to a normal distribution. The transformation can also be used for scaling down variables with large numerical values.

Multiplicative Transformation

In this transformation two variables are multiplied together to form a new variable because at times they affect the value of a property interactively. For example the variable quality might have a huge influence on the value of a property in a large, well built dwelling compared to a small dwelling constructed from cheap materials (Eckert, 1990). Multiplying the quality (QUAL) by the total living area (TLA) results in a new variable say QUALTLA that reflects the effect of each square meter of living area as a function of the construction quality. For example (Eckert, 1990);

If the coefficient of QUALTLA is R/m^2 200 and the scalar values for the variable quality are given as in table 3.1, the value per square metre for each class would be;

Excellent	= R/m^2 200x2.19	= R/m^2 438
Above average	= R/m^2 200x1.25	= R/m^2 250
Average	= R/m^2 200x1	= R/m^2 200
Below average	= R/m^2 200x0.77	= R/m^2 154
Poor	= R/m^2 200x0.22	= R/m^2 44

The above example shows that the QUALTLA variable will have a higher influence on the value of a dwelling with an excellent construction quality compared to one with an average construction quality even if they have the same total living area.

Quotient Transformation

The quotient transformation involves the division of one variable by another variable or by some constant value and is very helpful especially in situations where the range of values of one variable is large. For example during sales analysis, the price of the property cannot be used to assess the general trends within the property market because of the wide range of prices (i.e. from poor to affluent neighbourhoods) so the preferred indicator is the sale price per square metre (i.e. the sale price divided by the total living area). The advantage of the sale price per square metre is that it reflects the market relationships more accurately and prevents the model from being dominated by properties with very large sale prices (Eckert, 1990).

3.3 Time Adjustments

Due to variation in the sale prices of properties with regard to time of sale, they all have to be adjusted to the assessment date. They can be adjusted on a monthly, quarterly or yearly basis with the latter being imprecise since it reflects a general trend (Eckert, 1990). The rates (of inflation or deflation) can be computed based on a compound or constant (straight line) basis with the former being more precise and complex, but in the event of extreme inflation or deflation the two approaches produce the same results (Eckert, 1990). O'Connor (2002) states that time adjustments made prior to performing the regression analysis build an additional error to the estimated market value because smaller and lower valued properties tend to be overvalued and vice versa for higher value properties. It makes more sense to adjust the sale prices to the assessment date prior to modelling, especially when creating response surfaces, in that if it is built using sales that vary by large amounts due to different times of sale, then an error due to the difference in time should be expected in the estimated location factors. The four techniques used to generate time adjustment rates are (Eckert, 1990);

- Paired Sales Analysis
- Resale's Analysis
- Sales Ratio Trend Analysis
- Multiple Regression Analysis (MRA)

3.3.1 Paired Analysis

Paired analysis is based on the sales comparison approach, where two similar properties that have already been sold are compared. The value of the older property is adjusted to match the value of the newer property. The effects of the differences in property characteristics such as quality of building material and interior finish are first removed, and the remaining difference in value is attributed to inflation/deflation between sale dates. If a number of paired sales are available, the average inflation or deflation percentage can be derived, but the median is preferred because it is not susceptible to outliers. The equations 3.6 to 3.8(Eckert, 1990) show how the rates are calculated. The *value due to time* in equation 3.6 is equal to the sale price of the subject property less the amount attributable to the difference in characteristics of the two properties.

$$Rate/month = \frac{Value\ due\ to\ time}{Months} \qquad 3.6$$

$$Percent/month = \frac{Rate/Month}{Adjusted\ SP\ for\ Control} \qquad 3.7$$

$$Percent/year = Percent/Month \times 12 \qquad 3.8$$

Any sale price can then be adjusted using equation 3.9.

$$TAS = SP(1 + rt) \qquad 3.9$$

Where

TAS time adjusted sale

SP unadjusted sale price

r monthly or quarterly rate

t number of months from sale to assessment date

3.3.2 Resale Analysis

The resale analysis is similar to the paired analysis approach except that the same property is used for time adjustments, provided that a resale has occurred within a certain time period of the assessment date normally about two years. The new sale price is adjusted for any improvements that have been made since the old sale price. The major weakness of this approach is when there are an insufficient number of sales of this type to provide a reliable determination of the inflation rate.

3.3.3 Sales Ratio Trend Analysis

The ratio of the sale price to the estimated sale price derived from an initial MRA model without any time adjustments is determined and plotted against the sale month. The gradient of the best-fitting line through the data points, gives the percentage inflation or deflation. Eckert (1990) also states that this is a good technique for developing time adjustments (i.e. inflation or deflation).

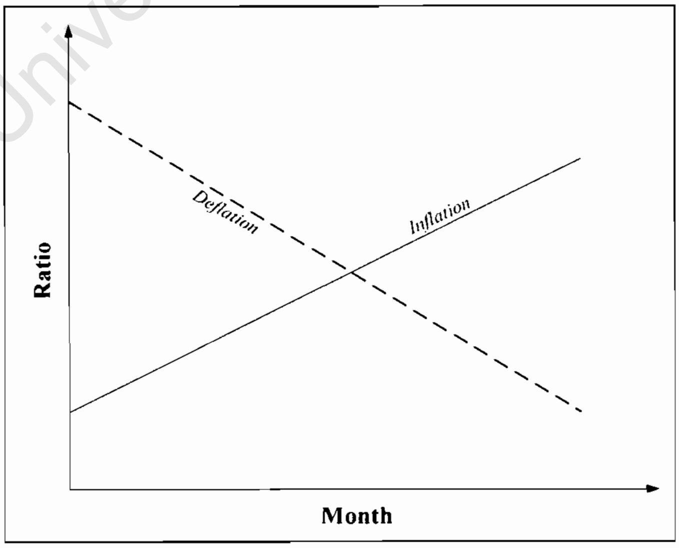


FIGURE 3.2 Inflation and Deflation Regression Lines

The month of sale is coded from the oldest *(1)* to the latest sale month *(n)*. Visual inspection of the gradient of the linear regression line gives an indication of whether there is inflation or deflation and Figure 3.2 illustrates this point. Equation 3.10 shows how the rate is computed and, using equation 3.9, the adjusted sale price can be determined.

$$Rate\ of\ increase/month = \frac{gradient}{intercept} = r \qquad 3.10$$

The advantages of this approach are that;

- The technique can utilize all available sales, whereas with the resale and paired analysis approach two similar properties must have sold within a certain time period.
- It is efficient, since no adjustments for physical or locational factors are required since there is no constraint of having to use similar properties.
- Market trends are readily seen on the graph.

In the event of a horizontal line arising then no adjustment needs to be made, but if a non-linear line emerges then this indicates a non-linear inflation/deflation, and finer analysis is required (analysis is divided into shorter time periods and summed). Adjustments factors have to be applied cumulatively as in equation 3.11.

$$\mathrm{TAS} = \mathrm{SP}(1 + r_1 \mathrm{t}_1 + r_2 \mathrm{t}_2) \qquad 3.11$$

3.3.4 Multiple Regression Analysis

Multiple Regression Analysis (MRA) attempts to quantify the influence of time, which is entered as an independent variable relating to the sale price. Dividing the coefficient of the time variable by the average sale price results in the rate or change per month and if multiplied by twelve months gives the inflation or deflation per year.

Assuming the average sale price is R200 000 and the coefficient for month of sale derived from the regression model is R600, the change per month can be computed as shown below.

$$Change/month = \frac{time\ coefficient}{average\ SP}$$

$$= \frac{600}{200000} = 0.003(0.3\%)$$

The inflation per year is equal to 3.6% (0.3%x12). The change per month can then be substituted into equation 3.9 resulting in a time-adjusted sale.

3.4 Model Building

A model is a representation of the relationship between various elements of a real situation. Models are used to test theories or predict outcome of events (Gloudemans, 1999). Models vary from being simple verbal or algebraic statements, to very complex mathematical equations. A model consists of a dependent variable, in this case the sale price (SP) of real property, and at least one independent variable, in this case the property characteristics. In CAMA, models are used to estimate the market value of real property. A simple model is represented by equation 3.12 and one that is representative of land and building factors by equation 3.13 (Eckert, 1990).

$$MV = IV + LV \tag{3.12}$$

$$MV = GQ[(BQ * \sum BA) + (LQ * \sum LA) + \sum OA] \tag{3.13}$$

Where

MV estimated market value

IV improvement values

LV land values

GQ general qualitative factors such as location

BQ building qualitative factors such quality of dwelling

BA building additive factors such as total living area

LQ land qualitative factors such as topography

LA land additive factors such as property extent

OA other additive factors such as swimming pools

Quantitative variables have an additive effect on the estimated market value, whereas qualitative variables have a multiplicative effect. The most critical component of CAMA models is data, because if they are developed based on wrong data then instability of models should be expected. The most effective way of checking model accuracy is through field review or applying the model to a test sample (Kochetkov, 1998).

In model development two processes are involved: model specification and model calibration. Model specification refers to the structure of the model that is what variables should be included and how they are functionally related to the estimated market value. Model calibration is the development of the coefficients of the independent variables through analysis of actual sales and property data (O'Connor, 2002).

3.4.1 Model Specification

Model specification is concerned with identifying all possible variables, testing their statistical significance in order to decide which variables should be included in the model. It is also used to identify variables that are highly correlated, and formulating a relationship between these variables and the dependant variable (estimate of market value). The property characteristics, from which the chosen set of variables are derived, are shown in table 3.2 and generally fall into one of the three basic groups:

- ***Building characteristics*** - these have to do with the main dwelling.
- ***Land characteristics*** - are related to the land on which the property is situated.
- ***Other characteristics*** – these are not directly related to the land or the main building.

TABLE 3.2 Property characteristics

Building	Land	Other
Overall Quality	Neighbourhood	Granny flat/ Servants quarters
General Condition	Property size/ ERF extent	Pool
Type of walls	Topography	Jacuzzi or hot tub
Dwelling type/design	Traffic/Street	Garage type
Age	View	Terrace/Balcony
Total living area	Garden	Tennis court
Common walls	Security	Shacks
Cooling or heating		
Number of rooms		
Number of bathrooms		
Interior/Exterior finish		
Roofing type		

Eckert (1990) states that reasonable models can be built using a total number of sales equal to or greater than four times the number of independent variables, while Isakson (2001) prefers that at least ten sales per variable be used. If a model is built with insufficient sales, either of the two scenarios stated below can be expected:

- The model can fail to solve for the coefficients of the independent variables especially if there are fewer sales than variables.
- A very unstable model can arise because it has been calibrated based on a very small sample, so when applied to a larger and varying group of properties it can fail to accurately predict values.

3.4.1.1 Types of Models

Additive Models

The additive model is represented by equation 3.14 and involves adding the property variables to a constant so as to enable the estimation of the market value. The function defines a linear relationship between the property characteristics and the estimated market value

$$MV = 150xView + 23xCond + 1000xTLA + 3562xTotfix + 245xErfext \qquad 3.14$$

The coefficients are the marginal contributions of each variable to the final market value and are determined in the model calibration phase. The problem with the model is that it fails to pick up the interactive relationships between variables (e.g. QUALTLA from multiplicative transformation) and can be inadequate in complex property markets.

Multiplicative Models

In multiplicative models the variables are multiplied together instead of being added, forming a non-linear relationship between the property characteristics and the estimated market value. In addition, the variables can be raised to powers or, as with binary variables, the variables themselves serve as powers as shown in equation 3.15.

$$MV = 1.2^{Villa} x0.9^{Court} xQual^{0.23} xView^{0.49} xTopo^{0.80} x(2396xTLA)x(131xErfext) \qquad 3.15$$

The binary variables are expressed as exponents because they can only take a value of one or zero that is, their effect is to either increase or decrease the value of the property and location is a typical variable that can be used as a binary variable. For example if using equation 3.15, the values 1.2 and 0.9 represent the coefficients for the neighbourhoods Villa and Court respectively. When estimating market values for properties from Villa, its binary value is made equal to one so that the property values are multiplied by the coefficient 1.2, whereas the binary value for Court is given a value of zero. The same process is conducted when dealing with properties from Court that is the binary value is equal to one, whereas that for Villa is given a value of zero so that it has no effect in the estimation of market values for properties from Court.

Multiplicative models capture the interactions present in the real estate market, but fail to separate land and building components so as to enable the actual effects due to each variable to be assessed on an independent basis.

Hybrid Models

A hybrid model is represented in equation 3.16, and is also referred to as a generic model. It incorporates both additive and multiplicative relationships between variables as seen in equation 3.16. Land variables (i.e. topography *(topo)*, property size *(erfext)* and security *(sec)*) and building variables (i.e. condition multiplied by quality *(CondQual)*, total living area *(TLA)*, total bathroom fixtures *(Totfix)* and external walls *(exwalls)*) are specified independent of each other. The variables that are assumed to have an overall (overall rate multipliers) effect on the property can be introduced as multiplicative factors and are location, street and view in equation 3.16. The variables that do not fall in either the land or building group can also be included in the model for example pool.

$$MV = Loc^{-0.17}\left[CondQual^{0.23}(2396xTLA + 497xTotfix) + Topo^{0.80}(131xErfext) + 152xPool\right] \qquad 3.16$$

The advantage of the hybrid approach is that it gives the modeller freedom to specify the model according to market conditions that might be existent at a particular time thereby resulting in more descriptive and higher quality models. Also the effect of each variable on the final market value of the property is easily explainable to both valuers and the general public. The only problem is that there is no unique solution when it comes to calibrating them since they exhibit a form that is neither totally linear nor multiplicative. They can be calibrated using any of the three methods: using non-linear MRA (trial and error process), multistage MRA and adaptive estimation procedure (AEP).

3.4.2 Model Calibration

Model calibration is the process of estimating regression coefficients in a mass appraisal model (Eckert, 1990). The estimated coefficients of the property variables provide the marginal prices that each variable in the model contributes to the total value of the property. Techniques available for calibrating sales comparison models are MRA and feedbacks. Additive and multiplicative models can be calibrated using the ordinary least squares method, which is the basis of MRA, but multiplicative models, are first transformed into a linear form by using the logarithmic transformation. Non-linear MRA and AEP are used for calibration of hybrid models.

AEP continuously adjusts the coefficients for each variable in the model based on the characteristics of each property encountered (Eckert, 1990). The adjustment process is repeated until all the sales in the dataset have been processed or the model converges on a satisfactory solution, therefore the term feedback. MRA attempts to minimize the sum of squared errors, whereas feedback seeks to minimize the average absolute error (Eckert, 1990). The major setback with AEP is that it does not provide statistical results such as R^2, t and f statistics, and other goodness of fit statistics, which come with MRA, but the results can be evaluated through the absolute percent error, which is similar to coefficient of dispersion. AEP has the advantage that it is less susceptible to outliers, thereby giving rise to stable results but great care must be taken when specifying the model, since it does not support stepwise regression. The ability for it to calibrate hybrid models gives it an edge over additive and multiplicative MRA (Eckert, 1990). As is the case with MRA, AEP requires skilled personnel to carry out the modelling because choice of variables and initial setting such as damping factors, number of iterations and initial weights is critical in achieving convergence and appropriate results.

3.5 Types of Valuation Techniques

Valuation methods can be grouped into either traditional or advanced, where in the former, the market value is arrived at through direct comparison with other properties or it is regressed as a function of the property and neighbourhood characteristics. In the advanced approach, simulation is used to try and replicate the thoughts of buyers and sellers (French et. al., 2003). Traditional methods include multiple regression analysis (MRA), comparable sales approach (CSA), income approach, cost approach and adaptive estimation procedure (AEP). The advanced methods comprise of Artificial Neural Networks (ANN), expert systems, spatial analysis, fuzzy logic and Autoregressive Integrated Moving Average (ARIMA). Application of a certain technique in a jurisdiction depends on available technology and the skills of the staff.

3.5.1 Advanced Techniques

3.5.1.1 Artificial Neural Networks (ANN)

Artificial Neural Networks are a form of artificial intelligence in that they try to replicate the way humans think. They consist of three basic components (French et. al., 2003):

- Input data layer
- Hidden layer referred to as the black box
- Output measure layer that consists of estimated property values

The ANN is first subjected to a process of training using sample data before it can be used for prediction purposes. The input data layer acts as a storage facility for the property characteristics before any processing takes place. The hidden layer is responsible for data processing and can be split into two components: weighted summation and transformation functions (French et. al., 2003). The weighted summation function is used to assign weights to the variables making up the input dataset. The transformation function is responsible for modelling a relationship between the summation values and the sale price thereby resulting in a network capable of estimating market values for unsold properties.

ANN have the ability to produce high quality results, but their disadvantage is that they are a form of a black box in that they just manipulate the data so as to come up with a network capable of estimating values, without taking into account the theory behind the data. The predictive accuracy of the network heavily relies on the initial selection of parameters for the network. Comparing ANN to MRA, it can be seen that a modeller dealing with MRA models needs a good understanding of the data otherwise the models will behave in an predictive manner especially if the variables are highly correlated with each. The advantage of ANN is that anyone can develop a predictive network without any prior knowledge of the data as the algorithm performs this task automatically.

3.5.1.2 Spatial Analysis Techniques

Spatial analysis techniques make use of three dimensional (3D) surfaces or grids interpolated from known points consisting of geographical coordinates *(x,y)* and a *z* variable which could be the sale price. They can be generated from interpolation techniques such as kriging or Inverse Distance Weighting (IDW). Figueroa (1999), with reference to Dubin (1992) and Pace *et al* (1998), says the likely problems to be encountered are:

- Multicollinearity between the independent variables.
- Very high and low values can be hidden because polynomials tend to smooth out such effects.
- Distortions are likely to occur especially if the data is unevenly distributed over a geographic area.

A more detailed discussion on spatial analysis techniques can be obtained in section 3.9.1.

3.5.2 Traditional Techniques

3.5.2.1 Comparable Sales Approach

This approach is based on the assumption that a buyer is well informed of how the property market operates and that the price paid for purchasing a certain property is a good estimate of the market value, and as such indicates the price of acquiring a comparable one in the same market (French et. al., 2003). The appraiser compares the property of interest to other similar sold properties; rarely will two properties be identical in terms of property characteristics hence the need to adjust the sale prices of the sold properties. Once the comparable properties prices have been adjusted they are then summed up and an average value determined, which corresponds to the market value estimate of the property of interest. The sales comparison approach is mainly used for residential property appraisals. The disadvantage of this approach is when comparable sales are very few, as a conclusive value might not be obtained.

3.5.2.2 Cost Approach

The cost approach assumes that a well-informed purchaser would pay no more for a property than the cost of purchasing the land and reconstructing the improvements thereon (Eckert, 1990). The appraiser estimates the land value only, but adds the cost of replacing the building and any other improvements. When the costs of constructing the improvements have been determined, the appraiser will account for physical, functional and economic depreciation so as to arrive at a reasonable market value. The cost approach is used when the property being appraised is relatively new because it has less depreciation and the construction costs are much easier to estimate. It can also be applied where there are an insufficient number of comparable sales. The cost approach is best applied to industrial and special purpose properties (e.g. steel plants), as they tend not to have sufficient sales (Eckert, 1990).

3.5.2.3 Income Approach

The income approach assumes that the value of a property is directly linked to the income it is expected to generate over its lifetime. It is valid in a market in which the owner is not the occupant and purchases the property for investment purposes that is for appraising income-generating properties. By analyzing the rental paid for the property and other comparables in the same geographic area the appraiser can then estimate the gross rental income of the property. Operating expenses are deducted from the gross rental so as to come up with a net operating income, which is then used to determine the value of the property. The value of the property is determined by capitalisation, where capitalisation is the ratio of the net operating income to the rate of return (Eckert, 1990). The rate of return is made up of three entities: discount rate, recapture rate and an effective tax rate.

3.5.2.4 Multiple Regression Analysis (MRA)

Multiple regression is another traditional approach, which is used to calibrate sale comparison models. Full details of this technique are covered in section 3.6.

3.6 Multiple Regression Analysis (MRA)

Multiple Regression Analysis (MRA) is a statistical technique for estimating market values using the sale price (dependent variable) and property characteristics (independent variable) (Eckert, 1990). Coefficients of the property variables are derived from sales analysis and reflect their individual contributions to the estimation of market values. Estimated values are accurate for properties with typical *(similar)* characteristics but a high margin of error should be expected for those with atypical characteristics. The aim of MRA in mass appraisal is to model the relationship between property characteristics and market value. MRA refers to a group of techniques for studying the straight-line relationships among two or more variables. Multiple regression estimates the regression coefficients b_i in the equation 3.17.

$$\hat{S} = b_0 + b_1 x_1 + \ldots + b_n x_n \qquad 3.17$$

Where the x_i represent the independent (predictor) variables and $\hat{S}$ is the estimated sale price. Although the regression problem may be solved by a number of techniques, the most used method is least squares. In least squares regression analysis, the b's are selected so as to minimize the sum of the squared errors in equation 3.18 between the sale price and estimated market value.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(S_i - \hat{S}\right)^2 \qquad 3.18$$

Where

$\hat{S}$ is the estimated market value of the property.

S is the sale price

e are the residuals

The smaller the value of $\sum_{i=1}^{n} e_i^2$, the better the regression analysis, it can be minimized by using more significant and non-correlated variables. It should be noted that the importance of any one variable in the regression equation is directly related to its contribution in reducing $\sum_{i=1}^{n} e_i^2$. The advantage of MRA over other traditional techniques is that it can handle more variables than the other techniques. The validity of MRA models depends on certain assumptions being met and these assumptions are covered in section 3.6.1.

3.6.1 MRA Assumptions

Violation of the assumptions stated below can lead to high instability of MRA models (Eckert, 1990).

Complete and Accurate Data

MRA models depend on data and therefore the estimated values cannot be any better than the data they are based on. Sale prices should be screened with non-arms length sales being removed and incomplete records should not be used in the modelling process.

Linearity

Linearity means that the marginal contribution (magnitude of regression coefficient) to the estimated value of an independent variable is constant over the entire range of the variable. The assumption merely means that a linear relationship exists between the variables. For example if the coefficient for the variable, size the of land is R/m^2 200, then all numeric values for the size of the land will be multiplied by R/m^2 200, even though in reality the value of land can increase up to a certain threshold there after the value starts to decrease (i.e. law of diminishing returns). Non-linearity is not a problem in multiple regression procedures provided that the standard deviation of the sale price is higher than the standard deviation of the residuals. A bivariate (two variables) scatter plot of the variables of interest can be used to assess linearity, if the deviation from this assumption is high, the variables can then be transformed using the logarithmic transformation. Also stratification of properties into homogenous groups can help to achieve linearity (Eckert, 1990).

Additivity

Additivity means that the marginal contribution of an independent variable in a model is not affected by the other independent variables, for example the assumption is violated when additional total living area adds more value in a new dwelling than an old dwelling (Eckert, 1990). Multiplicative and hybrid models are not affected by additivity, but in the event of the additivity assumption being violated by a certain variable then a quotient or multiplicative transformation can be applied to that particular variable.

Normally Distributed Errors

MRA assumes that residuals for all values of each independent variable are normally distributed. If this is not the case, the resulting *t* and *f* statistics are affected, and thus also the reliability of the individual regression coefficients. Usually non-normal residuals arise from poorly edited data or improper model specification. Violation of this assumption has no influence on the results obtained using a large dataset, but if the dataset is small then tests of normality have to be conducted and a typical test would be to produce a histogram plot of the data (Garson, 2003).

Homoscedasticity.

Homoscedasticity can be assumed to be a measure of quality of the estimated sale price to have a constant variance for all values of the property variables (Eckert, 1990). Homoscedasticity can be observed by making use of a plot of the regression residuals against any independent variable. If the cloud of points has a constant width, then homoscedasticity can be assumed to be present, whereas if the pattern has high variation in terms of width then violation of the assumption of homoscedasticity can be assumed and this is known as hetereoscedasticity. For example a funnel shape would imply that greater error should be expected as the sale price increases (Garson, 2003). Figure 3.3 shows the point clouds for homoscedasticity and hetereoscedasticity.

Hetereoscedasticity results in a model performing well within a certain range of prices and performing in an unpredictable manner outside this range. This can be overcome by using separate models for groups of homogenous properties. Alternatively using weighted least squares MRA or multiplicative regression, which has the ability to transform the dependent variable to follow a normal distribution using the logarithmic transformation.

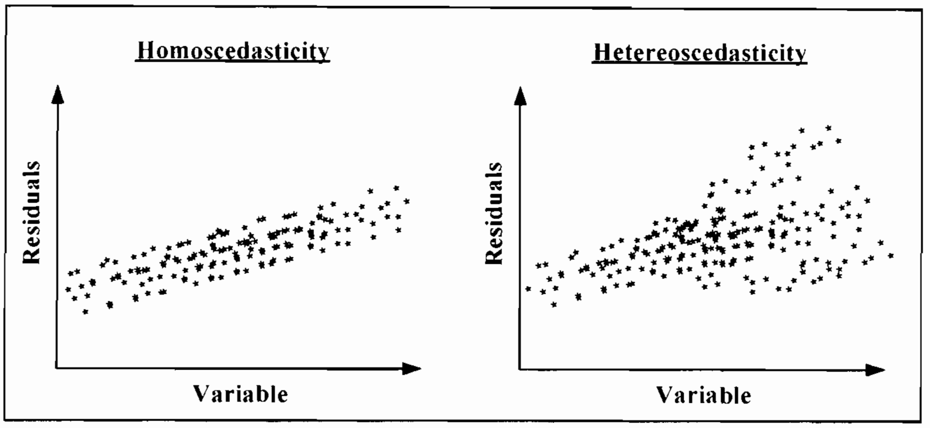


FIGURE 3.3 Point clouds

Uncorrelated Independent variables

If the independent variables are correlated they result in multicollinearity, which is very common in mass appraisal of real property. Two types of multicollinearity exist:

- Perfect multicollinearity, exists if there is a linear relationship or function between two or more variables, that is, one variable can be expressed as a constant multiple of the other.
- Imperfect multicollinearity occurs when some of the independent variables are correlated, but there is no exact linear relationship between them.

When perfect multicollinearity exists, the probability of obtaining a unique regression solution is very small, an example of such a situation is when there are fewer observations than variables. Imperfect multicollinearity results in a solution but the standard errors of the regression coefficients will be high, thereby reducing the level of reliability of the coefficients and assessment of the independent variables. Over fitting or use of redundant variables in an attempt to increase the value of the coefficient of determination *(R^2)* results in an increase in multicollinearity, therefore the need to omit redundant variables. Also use of the correlation matrix helps in identifying highly correlated variables so that they can be omitted from the model.

Representativeness

Representativeness refers to the closeness of the model design to the real situation. Model representativeness will decrease as the range of market factors increases (heterogeneity) and will increase with homogeneity. If a model is developed for high priced properties then it should be

- Failure to adjust or screen sales to remove personal property, which is not included in the definition of market value.
- Unusual property characteristics or a combination of characteristics.

In a large sample of accurate data, outliers would arise due to model weakness, for example multicollinearity, instability of complex functions and incorrect model structure.

3.7 Evaluation of Results and Ratio Studies

Quality of the results produced by the model can be assessed using the ratio of the estimated market value to the actual sale price. This ratio allows for the comparison between the estimated values to the actual sales prices, with a closer match indicating a high predictive accuracy of the model. The quality of results is affected by measurement errors, which are more common in qualitative characteristics. To avoid these pitfalls the appraiser should use a large representative dataset and conduct statistical tests for the models developed (Isakson, 2001). Eight statistics are used for assessing regression results and can be split into two groups: goodness of fit statistics and individual variable statistics (Eckert, 1990). Goodness of fit statistics are those used to assess the predictive accuracy or quality of the model and are the coefficient of determination *(R^2)*, standard error of estimate *(SEE)*, coefficient of variation *(CoV)* and the average percent error. The other group of statistics are used for assessing the significance of individual variables in the model and are the coefficient of correlation (r), the t-statistic, the F-statistic and the beta coefficient.

Coefficient of Multiple Determination (R^2)

The coefficient of multiple determination *(R^2)* is the percent of the variance in the sale price explained uniquely, or jointly, by the independent variables. The range of R^2 is from zero to one. When it is equal to zero it implies that none of the variation in the estimated market value is accounted for by the independent variables, whereas if the value is equal to one then all variation is due to the independent variables and the sum of the variances is zero that is $\sum_{i=1}^{n} e_i^2 = 0$.

Recalling from equation 3.18 that $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(S_i - \hat{S}\right)^2$

The sum of squared errors (SSE) is given by equation 3.19.

$$SSE = \sum_{i=1}^{n} (S_i - \overline{S})^2 \qquad 3.19$$

Where

S is the actual sale price

$\overline{S}$ is the mean sale price

The sum of squared deviations represented by equation 3.20, is the difference between the estimated market values and the mean sale price

$$SSE_{EX} = \sum_{i=1}^{n} (\hat{S}_i - \overline{S})^2 \qquad 3.20$$

Since the regression line cannot explain all the difference between the actual sales (S_i) and the mean sale price ($\overline{S}$), the sum of such remaining or unexplained squared errors, or residuals, (SSE_{UN}) is represented by equation 3.21.

$$SSE_{UN} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(S_i - \hat{S}_i\right)^2 \qquad 3.21$$

Therefore the sum of squared errors can be further represented by equation 3.22 and equation 3.23 shows how to compute the coefficient of multiple determination:

$$SSE = SSE_{EX} + SSE_{UN} = \sum_{i=1}^{n} \left(S_i - \hat{S}_i\right)^2 + \sum_{i=1}^{n} (\hat{S}_i - \overline{S})^2 \qquad 3.22$$

$$R^2 = \frac{SSE_{EX}}{SSE} = \frac{\sum_{i=1}^{n} (\hat{S}_i - \overline{S})^2}{\sum_{i=1}^{n} \left(S_i - \overline{S}_i\right)^2} \qquad 3.23$$

Shortcomings of R^2

a) As the number of variables increases so does the value of R^2 as long as they are correlated with sale price, but they may be insignificant therefore it is necessary to compute an adjusted R^2 ($\overline{R}^2$) as shown in equation 3.24.This new value, $\overline{R}^2$, is an adjustment for the fact that when one has a large number of independent variables, the value of R^2 can become artificially high because some independent variables will explain small parts of the variance of the estimated market value.

$$\overline{R}^2 = 1 - \frac{(n-1)SSE_{UN}}{(n-p-1)SSE} \qquad 3.24$$

Where

n is the sample size

p is the number of variables.

b) R^2 values should not be compared between two samples because of the differences in variance of the independent variables and dependent variable. For example an R^2 value derived for an entire jurisdiction should be interpreted differently to that generated for say three homogeneous areas.

Standard Error of the Estimate (SEE)

The standard error of the estimate is a measure of the amount of deviation between the sale price and the estimated market value that is the standard deviation of the regression residuals (Eckert, 1990). Equation 3.25 is used to compute the standard error of the estimate.

$$SEE = \left[\frac{\sum_{i=1}^{n} \left(S_i - \hat{S}_i\right)^2}{(n - p - 1)} \right]^{\frac{1}{2}} \qquad 3.25$$

From equation 3.25, the SEE can be interpreted as the square root of the sum of squared errors divided by the degrees of freedom and if the errors are normally distributed, then;

- 66% of sale prices will fall within one SEE of their estimated values
- 95% of sale prices will fall within two SEE of their estimated values
- 99% of sale prices will fall within three SEE of their estimated values

If the magnitude of the SEE is small (e.g. R2000) compared to the average sale price (e.g. R100000) the predictive accuracy of the model can be assumed to be good. If the average price is low (e.g. R20000) compared to a large value of the SEE of say R5000, then it can be concluded that the model is a poor predictor of value. Care should be taken when assessing model accuracy using this approach especially if dealing with properties that have extremely high and low values, as outliers will generally affect the mean sale price.

Coefficient of Variation (CoV)

The CoV is equivalent to the ratio SEE to the average sale price expressed as a percentage and is computed using equation 3.26.

$$CoV = \frac{SEE}{\bar{S}}.100 \qquad 3.26$$

Assuming the CoV has been computed to be 15%, it implies that about 66% of the sale prices will lie within 15% of the estimated market values. Similarly, 95% will lie within 30% of the estimated values.

Average Percent Error

Equation 3.27 shows the calculation that can be used for finding the average percent error, which is just the average absolute error between actual and predicted sales prices.

$$APE = \sum_{i=1}^{n} \frac{\left(\left|S_i - \hat{S}_i\right| / S_i\right)}{n}.100 \qquad 3.27$$

Coefficient of Dispersion (CoD)

The CoD is the amount of spread or variation from the typical ratio and is computed by dividing the average absolute deviation by the median ratio. The coefficient of dispersion (CoD) range that is expected is 5% to15%. Lower values under 5% are rare but can be achieved in homogenous jurisdictions, whereas heterogeneous areas will initially produce CoD values of up to 20% but should be reduced to 15% or less.

Coefficient of Correlation

Coefficient of correlation is a measure of the degree of linearity between two variables, and it ranges from +1 to −1. A correlation of zero does not mean that no relationship exists between the variables but that there is no linear relationship between them. A correlation matrix can be used to analyse the relationship that exists between two or more variables. If two independent variables are highly correlated, that is with a correlation coefficient close to one, then care must be taken in using both of them in the model, since they may behave unpredictably (i.e. multicollinearity).

Price Related Differential

The price related differential (PRD) is another statistical measure used to assess the uniformity of an assessment. It enables a check of whether higher and lower valued properties are assessed at the same level (vertical and horizontal equity). It is computed as seen in equation 3.28, by dividing the mean ratio (predicted to actual sale price) by the weighted mean ratio. The weighted mean is the ratio of the total market value estimates to the total sales prices of the entire sample.

It usually has an upward bias and should range between 0.98 and 1.03, except in a situation where they are a few sales (SDAT, 2001). A PRD above 1.03 indicates undervaluation of high priced properties relative to low priced properties (assessment regressivity). Assessment progressivity occurs when the PRD is below 0.98 and implies undervaluing of low priced properties relative to high priced properties.

$$PRD = \frac{\sum_{i=1}^{n} \frac{\hat{S}_i}{nS_i}}{\text{Weighted Mean}} \qquad 3.28$$

Where the Weighted Mean $= \dfrac{\sum_{i=1}^{n} \hat{S}_i}{\sum_{i=1}^{n} S_i}$

t-statistic

The t-statistic is used to assess the significance of the individual regression coefficients, b_i, specifically testing the null hypothesis that the regression coefficient is zero. The t-statistic can be calculated using equation 3.29.

$$t_i = \frac{b_i}{s_i} \qquad 3.29$$

Where s_i is the standard error of b_i (not SEE).

The t_i-*statistic* measures the marginal contribution of a variable in predicting the market value when all the variables are held constant. Although some variables maybe highly correlated with the sale price, they maybe insignificant as indicated by their t value because they duplicate information of other variables. A t- *statistic* in excess of ±2.00 gives an indication that one can be 95% confident that the independent variable is significant in the prediction of the market value and that in excess of ±2.58 indicates that one can be 99% confident. If the *t-statistic* is below the integer value of 2, it implies that the independent variable is insignificant in the estimation of market values.

F – statistic

The *f-statistic* is used to test the significance of R^2, which is basically the same as testing the significance of the regression model as a whole. If prob(F) < 0.05, then the model is considered significantly better than would be expected by chance and the null hypothesis of no linear relationship of the sale price to the independents is rejected. The *F-statistic* is a function of R^2, the number of independent variables *(p)*, and the number of sales in the dataset *(n)*, and can be computed using equation 3.30.

$$f = \frac{\left[\left(R^2 / p\right)\right]}{\left[\left(1 - R^2\right)/\left(n - p - 1\right)\right]} \qquad 3.30$$

Beta coefficients

Beta coefficients, β_i, are standardised regression coefficients that are useful for assessing the relative importance of the independent variables in a model, for example the age of property can be compared to the total living area variable because after standardisation they have the same units of measure. They can be computed by transforming the sale price and independent variables so that they have a mean of zero and variance of one (i.e. standardisation). Equation 3.31 shows how the beta coefficient is related to the regression coefficient.

$$\beta_i = b_i\left(\frac{s_i}{s_S}\right) \qquad 3.31$$

Where s_i and s_s are the standard deviations of the independent and dependent variables respectively.

3.7.1 IAAO Standard Ratios

The International Association of Assessing Officers (IAAO) has developed various standard ratios to gauge the quality of an assessment (IAAO, 1999). These are as shown in table 3.3. Models should be no more complex than necessary in order that they can be explainable to and understood by the property tax paying public, and not only the valuation profession. Coefficients should be generally consistent with appraisal theory. For example, size-related variables should be positive and age-related variables should be negative. It is every modeller's goal to have such an ideal model, but in reality not all models will succeed in achieving this goal. Modellers are of the opinion that if accuracy is to be achieved, then the model need not be as taxpayer friendly in explaining the estimated values, as more accurate models may well be more complex.

TABLE 3.3 Standard ratios

Type of Property	Measure of Central Tendency *(Mean and Median Ratios)*	CoD	PRD
Single family residential			
Newer, more homogenous areas	0.90 – 1.10	10.0 or less	0.98 – 1.03
Older, heterogeneous areas	0.90 – 1.10	15.0 or less	0.98 – 1.03
Rural residential seasonal	0.90 – 1.10	20.0 or less	0.98 – 1.03
Income – producing properties			
Larger, urban jurisdiction	0.90 – 1.10	15.0 or less	0.98 – 1.03
Smaller, rural jurisdiction	0.90 – 1.10	20.0 or less	0.98 – 1.03
Vacant land	0.90 – 1.10	20.0 or less	0.98 – 1.03
Other real and personal property	0.90 – 1.10	Varies with local conditions	0.98 – 1.03

The steps listed below can be used as a guideline to check the results produced by the CAMA models (Ward et. al., 2002):

- The coefficients produced by the model should be intuitive and the model must also be explainable.
- The sales ratio statistics should be within range of IAAO standards for the results to be assumed accurate and fair.
- The market values must be compared to the sales prices across the board so as to ensure equity.

The values generated by the model must be field-tested and also reviewed by a professional appraiser to check for consistency.

3.8 Stratification and Location Analysis

The large variation in property characteristics within a large and diverse metropolitan city brings up the need for them to be stratified. This is supported by the statement that from a mass valuation perception a predictive model requires the sub-division of a significantly large geographical area into realistic sub-markets or neighbourhoods (McCluskey, 2000). Sub-markets can be defined in a number of ways by employing the principle of stratification: a process of creating a number of homogeneous areas from a heterogeneous jurisdiction (McCluskey and Deddis, 2002). Stratification and location analysis can be accomplished in any of the three ways (Eckert, 1990);

Geographic Stratification

Geographic stratification involves boundaries being drawn along major economic or physical features such as mountains or major highways, so as to show the differences in location, resulting in separate models being specified for each area (Eckert, 1990). An economic area in this sense can be thought of as a grouping of neighbourhoods exhibiting the same economic forces or geographic location (Gloudemans, 1999). It is a good technique to employ if property values are homogeneous within neighbourhoods, but heterogeneous between neighbourhoods. Modelling using this approach is beneficial in that the sales used will reflect market influences and conditions only within that area, therefore enabling more accurate and defensible models to be produced.

Problems

- Multiple models give rise to inconsistencies in estimating market values when properties are reclassified into different neighbourhoods in successive valuation cycles.
- Contiguous properties in different neighbourhoods may have very different and unrealistic valuations and equity may suffer.
- Maintaining multiple models can be an administrative burden.
- If neighbourhoods are too homogeneous, calibration techniques cannot reliably estimate the effect on value of each characteristic, particularly those that occur infrequently.

Stratification based on Property Characteristics

Properties can be separated into relatively homogeneous strata or groups based on any property characteristic such as dwelling size. The modeller attempts to minimize the difference within a strata or group of properties whilst maximizing the difference between the different groups. Each group will have its own model developed. It is an appropriate approach when there is no distinct boundary between neighbourhoods and works best when the variation in property characteristics in a neighbourhood is more than between neighbourhoods (Eckert, 1990).

Single Model with Location Adjustments

An alternative to the two approaches stated above would be to use a single model that includes a location variable for the entire jurisdiction. The choice of model type (i.e. additive, multiplicative or hybrid) depends on the heterogeneity of the jurisdiction. Location factors for the model can be derived from location values response surfaces or incorporated in the form of delineated neighbourhood values. The single model approach has the advantage of reducing model building effort, but the major obstacle is that the model might not be able to predict market values that are representative of the heterogeneous jurisdiction.

3.8.1 Location Value Response Surfaces

Estimation techniques can be applied where there is a continuous measure being made at a particular location in space or time implying that the value at a location can be affected by its position and relation to its neighbours (Clark, 1979). The basic assumption with most techniques is that the relationship between the value of a point under consideration and any known value depends on the distance between them (Clark, 1979). Although the variable is spatially continuous, it is not always possible to take a sample of every location therefore the need to estimate unknown values from data taken at specific locations (Dorsel and La Breche, 1997). In spatial analysis the norm is to use a continuous variable with known geographic coordinates such as elevation to produce an estimate or model of the variable over the whole area, thereby resulting in a three dimensional (3D) surface.

The assumption made by most surface estimation algorithms is that the variable is continuous, unique and points at different locations are spatially autocorrelated. Working on the basis of these three factors, realistic estimates for the other points with unknown values of the variable in question can then be computed. Distribution of the points is crucial in that if they are widely spaced, there will be a lack of autocorrelation between them giving rise to a suspect surface. Therefore the spacing of points must be less than the distance at which the autocorrelation function declines to near zero.

Surface estimation techniques generally involve combining a method of selecting the set of points and a method of calculating the estimates from the known values. A number of methods can be used for selecting the points such as the neighbourhood criterion and triangulation. After identifying the points the estimates can be computed using a number of techniques such as inverse distance weighting or local polynomials. The success of surface interpolation techniques in generating a realistic surface heavily depends on the similarity of measured points and the techniques can be split into two major groups (Johnston et. al., 2001):

- ***Deterministic Techniques*** – generate a surface based on a mathematical function.
- ***Geostatistical Techniques*** – utilise both statistical and mathematical functions in surface interpolation and are capable of assessing prediction errors.

Choice of technique depends on the available resources such as software or computer processing speed.

3.8.1.1 Deterministic Interpolators

Deterministic interpolators have the ability to make the resulting surface pass through the original data and interpolators that are capable of achieving this task are known as exact interpolators, for example inverse distance weighting (IDW) and radial basis functions (RBF). This results in maximum and minimum values occurring at locations corresponding to known values. Inexact interpolators such as the global polynomial and local polynomial do not force the surface to pass through the known values therefore resulting in maximum and minimum values occurring at locations which do not correspond to the known values.

Deterministic interpolators can be further split into two groups:

- Global interpolators, they use the entire data to find unknown values for example IDW.
- Local interpolators, use neighbourhoods that are subsets of the entire dataset for example RBF.

Global Interpolators

Inverse Distance Weighting (IDW)

The IDW interpolator assumes that if objects are at close locations then it is highly likely that they will be similar. The technique estimates the value of the point of interest by weighting the influence of nearby data the most and more distant data the least. The estimate *(z')* can be computed using equation 3.32.

$$z' = \sum_{i=1}^{n} \lambda_i . z_i \qquad 3.32$$

Where the weight *(λ)* is determined as follows:

The distance between the prediction and known point locations *(d)* raised to a power *(p)* is be divided by the sum of all distances to produce the weighting factor *(λ)* as shown in equation 3.33.

$$\lambda_i = \frac{d_i^{-p}}{\sum_{i=1}^{n} d_i^{-p}} \qquad 3.33$$

The equation above implies that the weights will decrease exponentially as the distance increases. The weights are further constrained to sum to one that is

$$\sum_{i=1}^{n} \lambda_i = 1$$

The power *(p)* is determined by minimising the root mean square prediction error (RMSPE), which is a statistic quantifying the error of the prediction surface (Johnston et. al., 2001). If p=1 the function is a simple linear interpolator and when p=2, it becomes an inverse distance squared weighted interpolation, and is usually the preferred power. Figure 3.4 illustrates four different weighting functions.

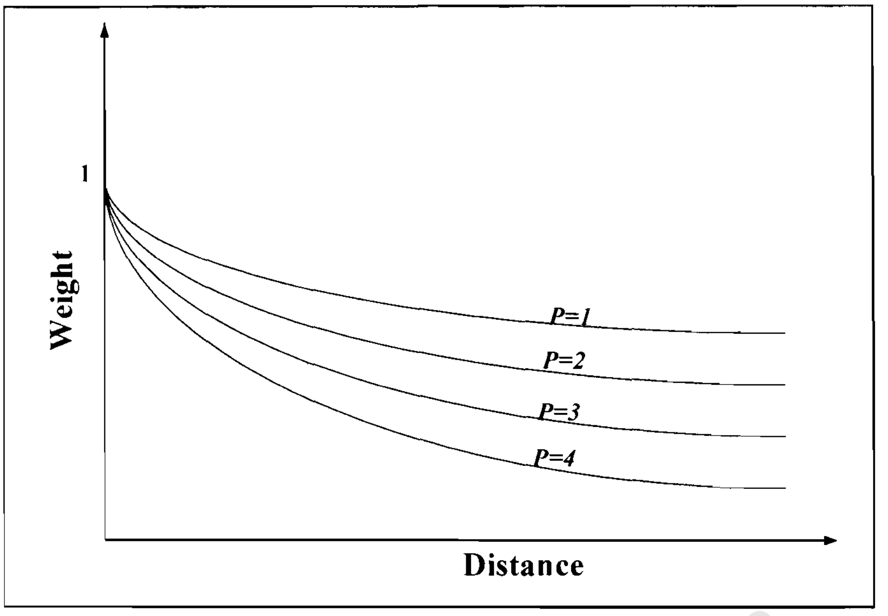


FIGURE 3.4 Effects of using different powers

The four functions can be seen to decrease with increasing distance, so the estimated value using p=4 will be derived from points very close to that location as compared to a case when p=1. Points to be used in the computation of the unknown value can be specified and this is known as the neighbourhood. The shape of the neighbourhood is influenced by the data, if there is a directional influence on the variable of interest then the best shape to use would be an ellipse, otherwise a circle is used (Johnston et. al., 2001). The major problem experienced with IDW is that it is sensitive to outliers and clustering of data points.

Splines

Splines are lines that pass smoothly and exactly through any number of given points in a sequential manner and are generally cubic in nature, although they can take higher powers their estimating ability at such high degrees is unpredictable. True curvature can be achieved by splines but their shortfall is that they do not add any value or improvement to the quality of the surface estimate (Swan and Sandilands, 1995). A splined surface is calculated from a fixed number of known points that are linked by, usually, a third order polynomial spline, thereby giving rise to a smooth surface with estimated values at all points. The third order polynomial is known as a cubic spline, it is a function whose first and second derivatives are continuous at the interpolation nodes.

When estimating values it should be expected to get a difference between the actual and estimate values, but the sum of the differences should be equal to zero - in other words they are unbiased. To fulfil the unbiased assumption the weights should be constrained to sum to one. Also the squared difference between the true and estimated values should be minimised by differentiating the function represented by equation 3.41.

$$\left(z' - \sum_{i=1}^{n} \lambda_i z_i\right)^2 \qquad 3.41$$

Minimisation results in the kriging equations 3.42 through to 3.44 (Johnston et. al., 2001).

$$\Gamma \times \lambda = g \qquad 3.42$$

$$\begin{bmatrix} \gamma_{11} & . & . & \gamma_{1n} & 1 \\ . & . & . & . & . \\ & & & & \\ . & . & . & . & . \\ \gamma_{n1} & . & . & \gamma_{nn} & 1 \\ 1 & . & . & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ . \\ \\ . \\ \lambda_n \\ K \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ . \\ \\ . \\ \gamma_{n0} \\ 1 \end{bmatrix} \qquad 3.43$$

The weights can then be computed by taking the inverse of the Γ matrix to the right hand side.

$$\lambda = \Gamma^{-1} * g \qquad 3.44$$

The Γ matrix contains the semivariances between all known locations, whereas the g matrix contains all semivariances between the unknown point and the known locations. λ is the matrix of the weights, which are to be computed and K is a Lagrange multiplier introduced so as to ensure that the minimum possible estimation error is obtained.

Semivariance (γ)

The semivariance is a measure of the spatial dependence between points and its magnitude varies with distance, and direction between points. A smaller distance results in a small semivariance and vice versa. The semivariance will increase until its value is equal to the variance of all points and at this stage will remain constant resulting in a flat region on the semivariogram called the sill. The distance from the unknown point to the sill is known as the range or span of the regionalized variable. Points lying within this range are related to each other and all known points contained in the neighbourhood, will be used in the estimation of the unknown point (Dorsel and La Breche, 1997). The semivariance has the same units as the regionalised variable.

The curved lines in figure 3.5 represent the different semivariance models used to compute the semivariance. The semivariance for regularly spaced points in one direction *(isotropy)* can be computed using equation 3.45.

$$\gamma_d = \frac{1}{2N_d}\sum_{i=1}^{N_d}(z_i - z_{i+h})^2 \qquad 3.45$$

Where

z_i is the value of the variable at location i

z_{i+h} is the value at another point

N_d *(number of separating distances)* = number of points - number of lags (d)

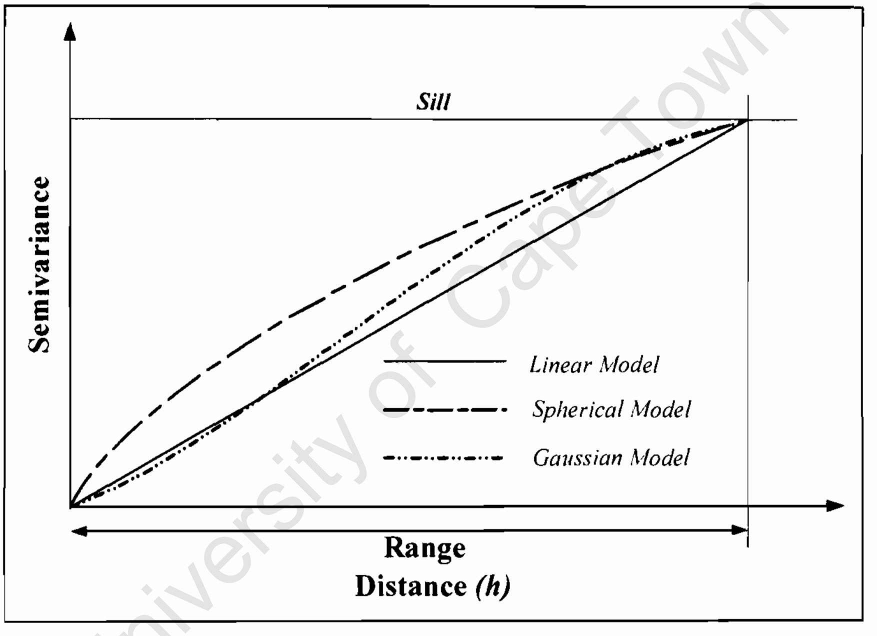


FIGURE 3.5 Semivariogram

The lag is a fixed distance interval. If the points are irregularly spaced then distance *(h)* becomes a vector instead of a scalar as in the isotropic situation. This case is known as anisotropy and uses the same equation as for the regularly spaced points. The semivariance is not restricted to being equal only to half the squared difference between pairs of points spaced at distance *h*, but it is also equal to the variance of the differences as seen in equation 3.46.

$$\gamma_d = \frac{1}{2N_d}\left[\sum_{i=1}^{N_d}\left((z_i - z_{i+h}) - \frac{\sum(z_i - z_{i+h})}{N_d}\right)^2\right] \qquad 3.46$$

The mean of z_i and z_{i+h} is the same since the same observations are used, but in a different order. This implies that their difference will be equal to zero as shown in equations 3.47 and 3.48 respectively.

$$\frac{\sum z_i}{N_d} = \frac{\sum z_{i+h}}{N_d} \qquad 3.47$$

$$\frac{\sum (z_i - z_{i+h})}{N_d} = \frac{\sum z_i}{N_d} - \frac{\sum z_{i+h}}{N_d} = 0 \qquad 3.48$$

As stated earlier the assumption is that the variable is stationary, if not the semivariance equation must be modified since the mean changes with distance.

The relationship between the semivariance and autocorrelation is as illustrated in figure 3.6.

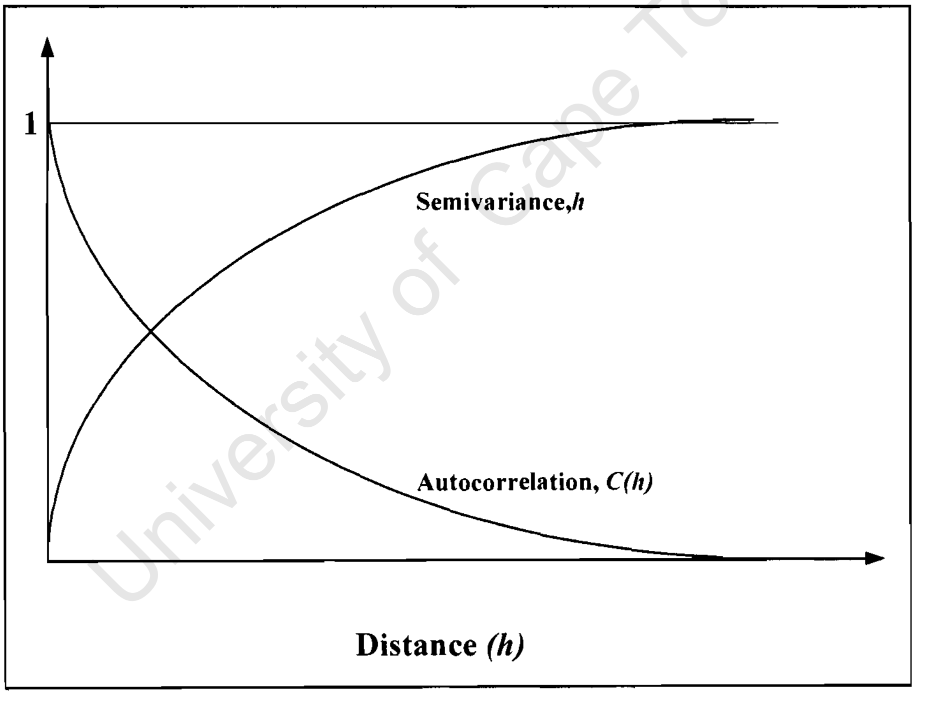


FIGURE 3.6 Semivariance vs Autocorrelation

Equations 3.49 and 3.50 represent the variogram and covariance respectively.

$$2\gamma_h = E(z_i - z_{i+h})^2 \qquad 3.49$$

$$C_h = E(z_i * z_{i+h}) \qquad 3.50$$

Expanding equation 3.49;

$$2\gamma_h = E\,(z_i - z_{i+h})^2$$
$$= E\,(z_i^2 + z_{i+h}^2 - 2 \times z_i \times z_{i+h})$$

$= E(z_i^2) + E(z_{i+h}^2) - 2 \times C_h$

$= 2\,(C_0 - C_h)$

$$\gamma_h = C_0 - C_h \qquad 3.51$$

Equation 3.51 shows the relationship that exists between the semivariance and covariance. C_0 and C_h are the covariances of the points at distances zero and h respectively.

Semivariogram Models

The literature on the different types of models expanded below is adopted from Chilè and Delfiner (1999).

Spherical Model

Spherical models show a linear form near the origin and reach the sill when the distance is equal to the range as seen in figure 3.7 and the model is represented by equation 3.52. It is the most used model because of a well-marked range and ease of calibration.

$$C_h = \begin{cases} 1 - \dfrac{3h}{2r} + \dfrac{h^3}{2r^3} & \text{for} \quad h \le r \\ 0 & \phantom{\text{for}} \quad h \ge r \end{cases} \qquad 3.52$$

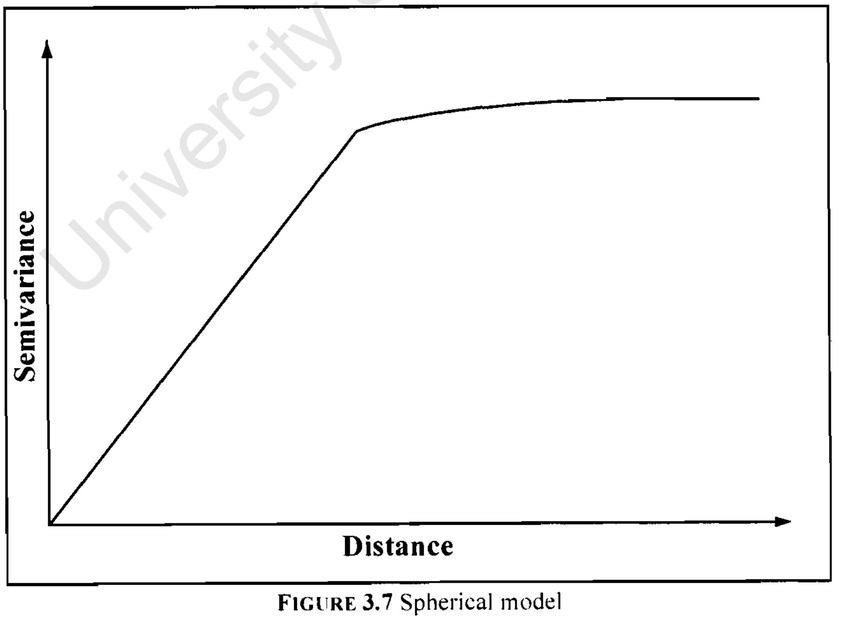


FIGURE 3.7 Spherical model

Exponential Model

The exponential model generally reaches the sill as the distance approaches infinity, but the practical range can be determined once the sill reaches about 95% of its value and this occurs at a distance equal to three times the range. Equation 3.53 and figure 3.8 represent the model.

$$C_h = e^{-\frac{h}{r}} \text{ for } r>0 \tag{3.53}$$

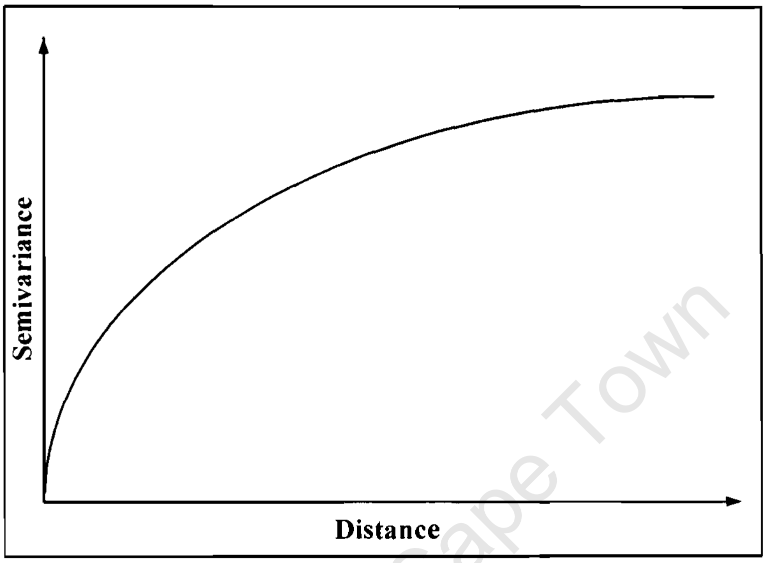


FIGURE 3.8 Exponential model

Cubic Model

The cubic model is used mainly where the variables are differentiable for example in areas such as pressure fields in meteorology. Equation 3.54and figure 3.9 represent the model.

$$C_h = \begin{cases} 1 - 7\frac{h^2}{r^2} + \frac{35h^3}{4r^3} - \frac{7h^5}{2r^5} + \frac{3h^7}{4r^7} & \text{for } h \le r \\ 0 & h \ge r \end{cases} \tag{3.54}$$

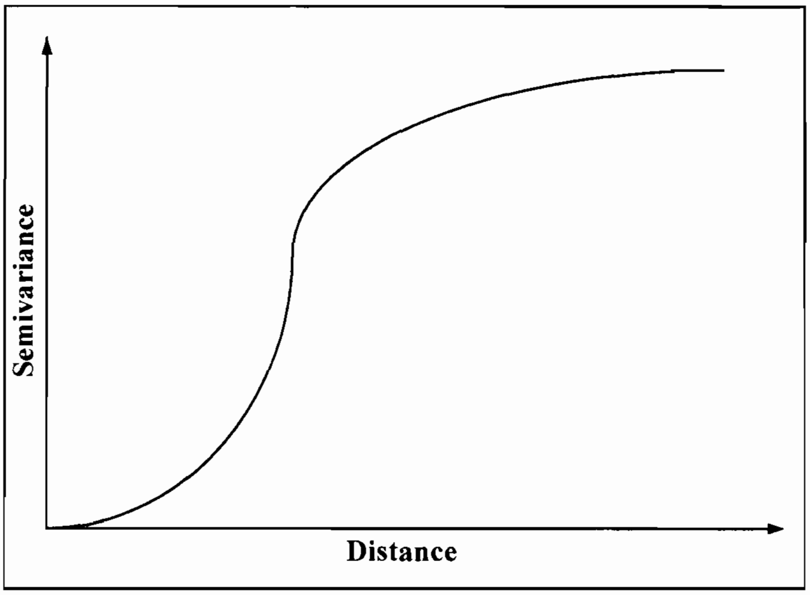


Figure 3.9 Cubic model

Gaussian Model

The practical range for this model is given as being around *1.73r*, with equation 3.55 and figure 3.10 representing the model, where r is the empirical range.

$$C_h = e^{-\frac{h^2}{r^2}} \text{ for } r>0 \qquad 3.55$$

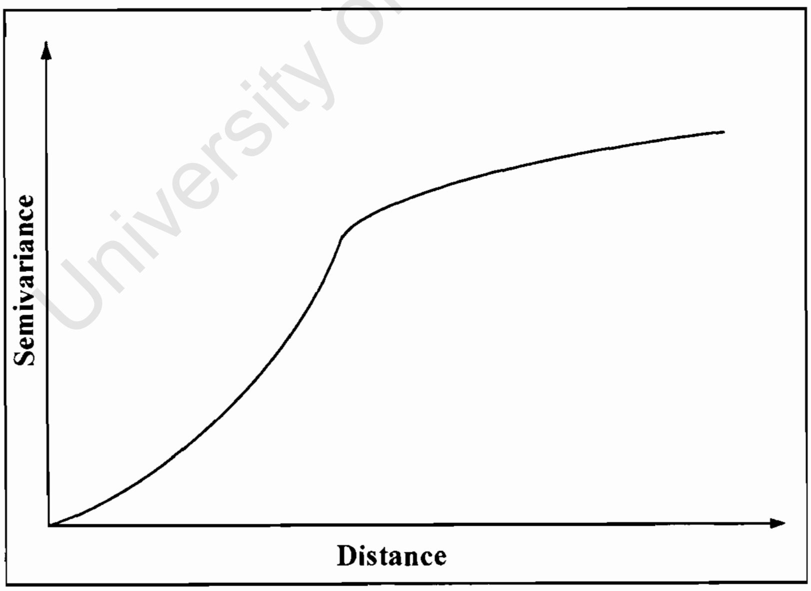


FIGURE 3.10 Gaussian model

The models are best fit onto the data by a geostatistician instead of automatically letting the software to fit it and this is due to fluctuations in the semivariogram. For manual fitting to take place effectively, sound knowledge about the data is crucial. If anisotropy is accounted for it will give rise to a complex but consistent and suitable global model. In automatic fitting, least squares is used, but it but will not produce satisfactory results.

Reliability of an estimate is measured using the variance of the kriged estimate, should it be small then the reliability can be assumed to be high. If the error is equal to zero then no local trend exists.

Problems of Kriging

When the regionalized variable is not stationary (the mean is not constant between samples) it is difficult to define a neighbourhood (an area encompassing the points to be used in the estimation) and therefore the weights. The solution to this problem is to split the variable into two entities that is the residual and drift. The residual can then be used as a regionalized variable thereby enabling a semivariogram to be produced. An arbitrary neighbourhood is chosen so as to allow for the drift to be computed. This creates another problem of selecting the correct semivariogram. The semivariogram for the residuals is assumed to be correct and is compared to a model semivariogram so as to allow for the estimation of the semivariance. If the two semivariograms are identical, the neighbourhood and semivariogram can be assumed to be correct, but if they differ then another semivariogram must be used. This is a highly interactive and subjective process.

Types of Kriging

Simple Kriging

Simple kriging assumes data is stationary, isotropic and has been collected at equally spaced points on a grid and estimation is for points only on the grid. It does not constrain the weights, that is they do not sum to unity and works with the residual from the mean. The basis of the method is that it tries to find weights, which it assigns to known values so as to estimate the unknown point. The weights are computed such that the effect of the relatively high variance of sample points is minimized by incorporating knowledge of the covariance between the estimated point and other sample points within the range. The estimate is arrived at using equation 3.56.

$$z' = \omega_1 x_1 + \omega_2 x_2 + \ldots\ldots + \omega_n x_n \qquad 3.56$$

Simple kriging uses the average of the entire data set and generally produces a result that is "smoother" and more aesthetically pleasing.

Punctual or Ordinary kriging

Punctual kriging uses the basic model given by equation 3.57.

$$z_i = \mu + \varepsilon_i \qquad 3.57$$

Where μ is an unknown constant mean.

It estimates the unknown value using the weighted values of nearby sampled points. The regionalized variable is assumed to be stationary and so no drift exists. Weights in ordinary kriging are constrained to equal to one, therefore the mean does not need to be known. It uses semivariogram information to estimate surfaces from known values. The estimated value can be assumed to be unbiased since no drift exists and the weights have been constrained to sum to one. The kriging variance can be used as a measure of the reliability of the estimate at each location. Punctual kriging is a more rigorous procedure when compared to simple kriging.

Universal Kriging

Universal Kriging is the most general form of operation in which values of blocks, areas, and points may be estimated from irregularly distributed samples in the presence of trends or non-stationary data. It uses the model represented by equation 3.58.

$$z_i = \mu + \varepsilon_i \qquad 3.58$$

μ is a deterministic function which, when subtracted from the data, leaves only the residuals and they are used to compute the surface. Revisiting equation 3.46, which represents the semivariance below:

$$\gamma_d = \frac{1}{2N_d}\left[\sum_{i=1}^{N_d}\left((z_i - z_{i+h}) - \frac{\sum(z_i - z_{i+h})}{N_d}\right)^2\right]$$

The equation is made up of two components: the difference between pairs of points, and the average of the differences. If the variable is stationary then the average of differences will be zero because the mean is constant. If not the variable can be assumed to be made up of the residual and the drift, the drift being the weighted average of points within a neighbourhood encompassing the unknown point. The residual is the difference between the variable and the drift. By removing the drift the residual can then be assumed to be a regionalised variable, thereby allowing for a semivariogram to be computed.

Universal kriging can be thought of as consisting of three steps

- Achieving stationarity (removing the drift or trend from the data)
- Kriging the stationary residuals
- Adding estimated residuals to the drift so as to enable computation of the estimates

Indicator Kriging

Uses the model given by equation 3.59.

$$I_i = \mu + \varepsilon_i \qquad 3.59$$

Where μ is an unknown constant and I_i is a binary variable. A variable that is continuous is first converted into a binary variable by choosing a threshold value: those values below the threshold level become zeros and above it become one (Johnston et. al., 2001). It uses the same principle as ordinary kriging.

Probability Kriging

The model used for probability kriging is given by equations 3.60 and 3.61.

$$I_i = I(z_i > c_i) = \mu_1 + \varepsilon_{i1} \qquad 3.60$$

$$z_i = \mu_2 + \varepsilon_{i2} \qquad 3.61$$

The means μ_1 and μ_2 are unknown constants and $I(z_i > c_i)$ is the threshold indicator that enables the creation of a binary variable I_i. Its operating principles are the same as indicator kriging but uses co-kriging to get better results. It attempts to exploit additional information from the original dataset as well as that due to the binary variables created. The major set back is that one needs to estimate the autocorrelation for each variable as well as their cross correlation and usually leads to more uncertainty being introduced (Johnston et. al., 2001).

Disjunctive Kriging

The model used is given by equation 3.62.

$$f(z_i) = \mu + \varepsilon_i \qquad 3.62$$

μ is an unknown constant and $f(z_i)$ is an arbitrary function. The output is an estimated value or indicator of the location with an unknown value. Good results can be achieved using this method but it also comes at a cost (Johnston et. al., 2001):

- Data must satisfy the bivariate normality assumption.
- Requires approximations for the functions $f_i(z_i)$ of which the functions are difficult to verify.
- Also the solutions are mathematically and computationally complex.

Co-kriging

At least two variables are used to predict the variable of interest by generating autocorrelations and cross correlations between the variable of interest and the other variables (Johnston et. al., 2001). The problems that are likely to be encountered are:

- More estimation has to be conducted.
- Variability in the autocorrelation is introduced when estimation is done for the unknown parameters.

3.8.1.3 Trend Surface Analysis versus Kriging

A kriged estimate is based on adjacent points and distant points have no influence on the determination of the unknown value, whereas in trend analysis, all the points are used in estimating the unknown value. Kriged estimates tend to be more stable than those due to trend surface polynomials, when estimated from the same number of data points. If the points are sparsely distributed, the trend surface polynomial tends to produce estimates that are suspect, which is unlikely with kriging techniques. Simple kriging produces poor estimates compared to polynomials. Trend surface techniques have the advantage of enabling tests of significance since they are based on multiple regression, which is not feasible with kriging functions. The methods used in kriging allow it to have an advantage over other estimation procedures in that the estimated values have a minimum error associated with them and this error is quantifiable (i.e. can generate an error surface).

3.8.1.4 Surface Errors

The three errors that are most likely to occur on any surface (Clark, 1979) are:

Edge Effects

An area is defined by boundaries and most a time they is no data at the edges or the map area extends beyond the known value points thereby resulting in unrealistic estimates at the edges and these are known as edge effects. If working with a polynomial function, it will assume the gradient from the last known point to the edge to be the same, which in reality maybe not be the case.

Bulls Eye Effect

Contours circling data points because they are the locus of high and low values, so some techniques will produce estimates that do not exceed the range of the known values, although higher unknown values can exist between the known points. This is very common with splines.

Honouring Known Points

Depending on the technique used, some will produce surfaces giving estimates at control points which are equal to the known values such as splines and inverse distance weighting. Some, however, result in estimated values at points differing from the known values.

3.9 Software

Choice of software depends on (Ward et. al., 2002):

- Knowledge of available packages that employ both statistical and spatial techniques.
- Costs.
- Size of jurisdiction.
- Accuracy required (the sophisticated the package, the higher the quality of results)
- Integration and connectivity to other systems.

In the short term investing in a cheap package is okay, but if it is for the long term then it is worth going for the sophisticated systems.

3.9.1 Comparison between NCSS and SPSS

Number Cruncher Statistical Systems (NCSS) and Statistical Package for the Social Sciences (SPSS) are the common packages used for computer assisted mass appraisal (CAMA) modelling. The two statistical packages can safely be assumed to deliver similar outputs, as all statistical functions running in the background are identical. For example, a sample mean of a dataset will always be the same regardless of whether a calculator or statistical package is used to compute it. The major difference that exists between the packages can be attributed to the ease or difficulty of operation. Of the two packages encountered to date, that is (SPSS) and (NCSS), three major differences can be identified (Ward, 2004).

Data Handling Capabilities

NCSS operates in two modes: as a spreadsheet (S0) and database (S0Z). If the total number of records in a dataset is below 16 000, it is stored in the spreadsheet section, but if it exceeds this number, the data has to be transferred to the database section. The processing speed drops as the number of sales approach 16000 and the situation becomes worse when the database mode is used. The amount of data SPSS can accommodate is unlimited and the processing speed is very high even when working with a large dataset. SPSS also handles alphanumeric data easily but this benefit is offset by the fact that working with the output (results) can be cumbersome as additional editing skills are a necessity. The opposite is true for NCSS with regard to these two aspects.

Costs

The purchase cost of SPSS is relatively expensive, with a single user license costing as much as US$3000, whereas NCSS costs $400.

Regression Analysis

The NCSS regression module is better, especially with regard to non-linear regression where variables are specified through a template. SPSS, however, requires that an equation representing the sales comparison model be typed in. The SPSS approach can be very "dangerous" especially if an error is made whilst typing. An example would be missing out or putting a comma in the wrong position in the syntax of the model which can throw everything into disarray. A high level of concentration is therefore required compared to NCSS, whose use of templates reduces such errors.

NCSS has a slight advantage over SPSS because of its ability to generate location value response surfaces, which can be an advantage if spatial analysis packages such as ArcGIS are not available.

3.9.2 Number Cruncher Statistical Systems (NCSS) Commands

Descriptive Statistics Module

It is a procedure for summarizing variables both statistically (mean, median and measures of spread) and graphically.

Mass Appraisal Module

This procedure enables the modeller to specify and calibrate the model.

Linear Regression Module

The procedure enables a straight-line relationship between two variables to be formulated thereby resulting in a linear equation and an error term.

LookUp Command

Example of view: Lookup *(View, View code, Scalar value of view, Compare)* returns the scalar value of view that is on the same row as the value of the view code that matches the current value of view.

Correlation Types

- **Pearson Product-Moment** – the most widely used measure of correlation and calculated, by summing up the products of the deviations of the scores from the mean.
- **Spearman Rank** – used for ranked *(relative)* data, rather than *linear* measurements and preferred when Pearson's correlation coefficient is undesirable or misleading.

Stepwise Regression Types

- **Forward, or Step-Up** – This method is often used to provide an initial screening of the candidate variables especially when dealing with a large group of variables.
- **Backward, or Step-Down** – This method builds a model with all variables in the dataset thereby making it less popular because may include variables that are not really necessary thereby resulting in a large R-Squared value.
- **Stepwise regression** – is a combination of the forward and backward selection techniques but mainly a modification of the forward selection. When a variable is added the other variables in the model are checked to see if their significance has been reduced below the specified tolerance level of which they are removed from the model.

3.9.3 ArcGIS Commands

Geostatistical Extension Module

This module is used for advanced surface modelling using deterministic and geostatistical techniques. It has a wizard, which enables surfaces to be generated using Radial Basis Functions (RBF), IDW, Local, Global and Kriging interpolators.

Histogram Tool

The histogram tool enables a univariate (one-variable) description of the dataset to be examined as it is capable of displaying the frequency distribution (graph) and also calculating the summary statistics (mean, median and standard deviation).

Normal QQ Plot

It is a graph used for displaying the values of two distributions and if the data is normally distributed, it will lie on a straight line but will deviate from a straight line if not.

Trend Analysis

Used for mapping trends that may exist within the dataset in a three dimensional perspective.

Transformations

These transformation are applied when they is a need to change the dataset so that it follows a normal distribution. The types of transformations that can be used are:

- **Log Transformation** – often used where the data follows a skewed distribution.
- **Box Cox (Power) Transformation** – used when the differences or variation within dataset varies is small for certain areas and large for another.
- **Arcsine Transformation** – used for data that is represented as proportions or percentages.

Extent Tool

Used to remove excessive information for example when a surface is created it will overlap the area of interest and this tool enables the surface to be resized so it fits only the rectangular area of interest only.

Clip Tool

Used to create new layers that are subsets of a larger layer.

Manual Method

The modeller specifies the intervals manually.

3.9.3 ArcGIS Commands

a. *Geostatistical Extension Module*

This module is used for advanced surface modelling using deterministic and geostatistical techniques. It has a wizard, which enables surfaces to be generated using Radial Basis Functions (RBF), IDW, Local, Global and Kriging interpolators.

b. *Histogram Tool*

The histogram tool enables a univariate *(one-variable)* description of the dataset to be examined as it is capable of displaying the frequency distribution *(graph)* and also calculating the summary statistics *(mean, median and standard deviation).*

c. *Normal QQ Plot*

It is a graph used for displaying the values of two distributions and if the data is normally distributed, it will lie on a straight line but will deviate from a straight line if not.

d. *Trend Analysis*

Used for mapping trends that may exist within the dataset in a three dimensional perspective.

e. *Transformations*

These transformation are applied when they is a need to change the dataset so that it follows a normal distribution. The types of transformations that can be used are:

i. **Log Transformation** – often used where the data follows a skewed distribution.

ii. **Box Cox (Power) Transformation** – used when the differences or variation within dataset varies is small for certain areas and large for another.

iii. **Arcsine Transformation** – used for data that is represented as proportions or percentages.

f. *Extent Tool*

Used to remove excessive information for example when a surface is created it will overlap the area of interest and this tool enables the surface to be resized so it fits only the rectangular area of interest only.

g. *Clip Tool*

Used to create new layers that are subsets of a larger layer.

h. *Manual Method*

The modeller specifies the intervals manually.

Chapter Four

CAMA and Response Surface Modelling

This chapter is mainly concerned with data exploration and model development. Data cleaning is the first process, with records that are non-arms length or assumed to be in error being identified and removed. Stratification follows with the records being split into two samples: the test sample and modelling sample. The modelling sample is adjusted for time so as to reduce the sale prices to the assessment date, thereby enabling the last process in data analysis, the linearisation of qualitative data to be carried out. The modelling sample is then used to develop three base models without the inclusion of any location effects. The first model consists of properties for the entire Cape Town MLC, whilst the second model consists of properties from the around Table Mountain area. The third model is made up of properties from the Cape Flats and Southern suburbs. Three base models are used, as they will enable the effect (i.e. predictive accuracy) of developing models based on properties that are either very heterogeneous or slightly homogenous to be quantified. Location factors derived from different interpolation techniques are then incorporated into the base models. The processes conducted on the modelling sample are then applied to the test sample so as to enable checks for validity, and reliability to be conducted.

4.1 Exploratory Data Analysis

4.1.1 Data

The data to be used in this research is provided by the Cape Town City Council and consists of all single property residential sales, for the period January 1999 to December 2000, used in the General Valuation 2000 (GV2000) Project. The file consists of 24 912 records for the Cape Metropolitan Area (CMA), or Unicity, and the CMA consists of six former Metropolitan Local Councils (MLC):

- Cape Town
- Blaauwberg
- South Peninsula
- Tygerberg

- Oostenberg
- Helderberg

The property records of interest to the research are only those for the Cape Town MLC. The research is conducted for the Cape Town MLC only, because if it is carried out for the entire CMA, conclusive results might not be obtained due to time constraints imposed on the research. The research is carried out using the experimental design based on independent measures. Independent measures in the sense that each property record is used only once for either modelling or testing purposes and not both. The dependent variable is the sale price, which is the price of the property recorded at the Register of Deeds or on the Rates Clearance Certificate (Ward, Ireland, Field and Epstein, 2000). The independent variables, are the property characteristics and are as stated in table 4.1, but the original form used in the GV2000 exercise is contained in *Appendix A*.

4.1.2 Hardware and Software

For data processing and manipulation, an Intel(R) Pentium 4 computer with a processing speed of 2.00GHz and RAM of 512Mb running on the Microsoft Windows 2000 operating system is used in this research. A number of software packages are used for data manipulation and are:

- *Microsoft Excel* – spreadsheet used for data cleaning and linearisation.
- *Microsoft Access*– used for creation of new datasets.
- *ArcGIS 8.3*– for visualisation and spatial analysis.
- *Number Cruncher Statistical Systems (NCSS)* – statistical package used for statistical analysis and modelling.

TABLE 4.1 Property characteristics

Variable	Acronym	Remark
Street	Str	Type of street the property is situated on
Parking	Park	Availability of on or off street parking
Security	Sec	Level of security on dwelling
Topography	Topo	Nature of terrain e.g. sloping or flat
View	View	Type of view e.g. ocean or mountain
Garden	Gard	Whether well or poorly maintained
Dwelling Type	Dtype	Design of dwelling
Common Wall	Cwall	Detached or attached dwelling
Condition	Cond	Physical condition of the dwelling
Quality	Qual	Quality of material used to construct the dwelling
Pool	Pool	The size of pool if available
Total Living Area	TLA	Total living area of the dwelling in metres squared
Total Fixtures	Totfix	Total number of baths, showers, basins and toilets
Granny Flat	Grfl	Secondary dwelling
Garage or Carport	Garcp	Closed or open structure for parking vehicles
Storeys	Storeys	Number of stories in the main dwelling
Roofing	Roofing	Type of roofing cover
External Walls	Exwalls	Type of material used to make the outside wall
ERF Extent	Erfext	Size of the lot or property

4.1.3 Data Cleaning

All property records are imported into *Excel* prior to the sales cleaning process, the purpose being to remove records that have missing or wrongly coded data. The shaded boxes in table 4.2, denote missing data and miscoded data for example SALMON (cumulative sale month) is given as 8 but in actual fact it should be 23 (11+12), where 11 represents the 11^{th} month (i.e. November) in 2000 and 12 represents the 12 months for the year 1999.

TABLE 4.2 Inconsistent records

PID	YR	MTH	SALMON	STR	PARK	SEC	TOPO	VIEW	GARD	DATE
CCT0001676	2000	2	14	2	2		1	3	3	2000/02/07
CCT0001681	2000	1	13							2000/01/27
CCT0001679			3							
CCT0001854	1999	6	6	4	2	1	1	1	3	1999/06/17
CCT0001690	1999	3	3	4	2	2	3	3	3	1999/03/09
CCT0000261	2000	11	8	4	1	3	3	5	4	2000/11/07
CCT0000491	1999	8	8	4			3	1		1999/08/30

After all records satisfying the above criteria have been removed, the next step is to eliminate sales assumed to be non-arms length or outliers. To achieve this, descriptive statistics *(mean, median and standard deviation)* for the sale price per total living area (SPSM) and sale price

(SP) for each neighbourhood *(see Appendix B)* are computed using the *descriptive statistics module* in *NCSS*. A column named *good* is inserted into the worksheet to denote a sales validity; a bad sale is assigned a value of three, and a valid sale is assigned a value of one. The neighbourhood the property is located in is used in the assessment of validity of the sale price. If the property's SPSM is greater than two or three standard deviations from the mean SPSM for the neighbourhood, it is labelled with a value of three implying that it needs to be further investigated for validity. If it lies within the required range then it is given an indicator of one, which subsequently enables it to be used in the modelling process. All vacant plots (i.e. with a total living area equal to zero) are labelled with a value of three. After the above processes have been conducted on all records, the valid sales are then saved in a different file named *Cape Town Cleaned*.

4.1.4 Stratification

The cleaned records are split into two files, the modelling sample (80% of the records) and the test sample (20% of the records) using the process of stratification. The modelling sample is used for base modelling, whereas the test sample is used for conducting validity and reliability checks. Each neighbourhood contributes data to each sample according to the respective percentages *(see Appendix B)*. The stratified sampling method is used to select records from each neighbourhood based on the sale price, for example in table 4.3 there are nine properties in Zonnebloem, so based on the 80:20 ratio the modelling sample will have seven properties and the test sample will have two. The properties are first sorted in ascending order off their sale price and the two properties are chosen using the stratified approach that is the first and third quartile properties are picked. The remaining seven properties are assigned to the modelling sample. If they are say hundred properties in a neighbourhood, applying the stratified sampling method, it implies that every fifth property should be assigned to the test sample when using the ratio 80:20. Therefore the modelling sample will have eighty properties, whereas the remaining twenty will be for the test sample.

TABLE 4.3 Stratified sampling

PID	YR	TOPO	VIEW	DATE	SALE	X	Y	NBHD
CCT0000803	1999	3	2	1999/11/11	198000	-52698.7	-3756444	ZONNEBLOEM
CCT0000781	1999	1	2	1999/09/29	210000	-52787.2	-3756403	ZONNEBLOEM
CCT0000823	1999	3	2	1999/02/09	216000	-52639.2	-3756476	ZONNEBLOEM
CCT0000624	1999	1	3	1999/09/20	248000	-52961.2	-3756046	ZONNEBLOEM
CCT0000601	1999	1	2	1999/04/17	279000	-52307.7	-3755978	ZONNEBLOEM
CCT0000785	1999	3	2	1999/08/25	280000	-52176.1	-3756422	ZONNEBLOEM
CCT0000798	1999	2	2	1999/07/03	280000	-52135.9	-3756440	ZONNEBLOEM
CCT0000604	2000	1	2	2000/02/22	290000	-52296.5	-3755990	ZONNEBLOEM
CCT0000783	1999	3	2	1999/03/17	329000	-52180.9	-3756420	ZONNEBLOEM

The modelling sample is then imported into *NCSS* to allow for time adjustments to be conducted on the records.

4.1.5 Adjustment for Time

The sales ratio trend analysis method is used to adjust the sale price of the properties to the assessment date (January 2000). A very basic model is first specified using the *mass appraisal module* in *NCSS*, so as to enable the generation of appraisal values without any time adjustments. The sale appraisal ratio for each property is derived through division of the sale price *(SP)* by the appraisal value (*AV*). A mean sale appraisal ratio is calculated for each month and using the *linear regression module,* a plot of the mean sale appraisal ratio versus month is generated. From this plot, a linear equation with the gradient representing the increase per month can be generated. The rate of increase or inflation per month (r) is obtained through division of the gradient by the intercept. The time adjustment function (equation 4.1) is derived and applied to sales within the interval January 1999 to December 2000 resulting in a time adjusted sale (TAS).

$$TAS = S.P[1 + r(13 - \{(YR = 1999)(Month) + (YR = 2000)(Month + 12)\})] \qquad 4.1$$

To carryout the modelling process using all the variables some have to undergo transformations so as to make them temporally compatible.

4.1.6 Linearisation

Microsoft Excel is used to perform the task of transforming the qualitative variables to scalar (linear) quantities using the mean and median values of the SP, and SPSM *(see Appendix C)*. The average class for each qualitative variable is identified and assuming a linear scale is in use, then its value based on that scale should be one. Equation 4.2 (Ward, Ireland, Field and Epstein, 2000) is applied to all qualitative variables thereby resulting in scalar values. The derivation of scalar values need not be restricted to equation 4.2, as any method can be used. Using the sale price alone poses the problem of the mean being affected by outliers, while the SPSM is affected by the size of the property that is a large property tends to have a small SPSM, whereas a very small property will have a high SPSM. For example a $100m^2$ property selling for R200 000 will have a SPSM of $R/m^2 2000$, whereas a $200m^2$ property selling for R300 000 will have a SPSM of $R/m^2 1500$. Equation 4.2 has an edge over using the sale price and SPSM separately as it takes into account both factors.

$$\text{Scalar Value} = \sqrt{\left(\left(\left(\frac{\text{Median SP of each class}}{\text{Median SP of Average Class}}\right) + \left(\frac{\text{Median SPSM of each class}}{\text{Median SPSM of Average Class}}\right) \times 2\right) / 3\right)} \qquad 4.2$$

After transformation of the qualitative variables, the class codes and their corresponding scalar values for all the qualitative variables are copied from *Excel* to *NCSS*. The scalar values are assigned to the entire dataset through the *LOOKUP* command in the variables section *(see Appendix D)*, therefore paving the way for the modelling process.

4.2 Modelling

A correlation matrix using *Pearson Product Moment* in the *correlation module* is first run. All property characteristics are used in generating the matrix *(see Appendix E)* with the sale price being used as the response variable. A backward stepwise regression, which is found in the *regression module*, is run so as to determine the most likely influential variables *(see Appendix G)*. The response variable is set to the sale price, whereas the independent variables consist of the property characteristics. With an idea of the most important variables the process of creating base models is then initiated.

4.2.1 Base Model

The building, overall (i.e. their effect is to either increase or decrease the estimate) and land variables are specified in the *mass appraisal module*, resulting in a hybrid base model without locational adjustments. Three base models are specified as shown in table 4.4.

TABLE 4.4 Model specifications

Model	Overall	Building		Land		Other
		Qualitative	Quantitative	Qualitative	Quantitative	
CT	Park, View	Cond_Qual, Storeys, Cwall, Roofing, Dtype, Exwalls	TLA, TotFix	Topo, Sec	Erfext	
ATM	Str, View	Cond_Qual, Exwalls	TLA, TotFix	Topo, Sec	Erfext	Pool
NATM	Park, View	Cond_Qual, Dtype, Cwall, Exwalls	TLA, TotFix	Str, Sec	Erfext	Gard, Pool, Garcp

Where

CT consists of neighbourhoods within Cape Town MLC

ATM consists of neighbourhoods around Table Mountain

NATM consists of neighbourhoods not around Table Mountain

The neighbourhoods falling within each category can be seen in *Appendix H.* Calibration of the models is achieved by running the *mass appraisal module.* These models are then used for incorporating location factors generated by the different location analysis techniques.

4.2.2 Location Analysis

The location factor is derived from the *zscore* (Ward, Ireland, Field and Epstein, 2000), which is just a standardised value for any variable. *Zscores* have an advantage of compatibility, since they are just standardised variables. In the event of insufficient sales, *zscores* derived for vacant properties and those deduced from improvements on the property can be combined for surface generation purposes. Another advantage is the issue of scale. The magnitude of the *zscore* is smaller than that of the SPSM, and this results in faster processing times especially during surface generation. Also using the SPSM for generating location factors result in problems such as multicollinearity because the sale price is used as a dependent variable at the same time as an independent variable. Two *zscores* are computed, one based on the sale price per square metre *(zscore_SPSM)* (equation 4.3) and the other on property characteristics *(zscore_char)* (equation 4.4).

$$zscore\_SPSM = \frac{SPSM - \mu_{SPSM}}{\sigma_{SPSM}} \qquad 4.3$$

$$zscore\_char = \frac{x - \mu_x}{\sigma_x} \qquad 4.4$$

Where

x property characteristic

μ_x mean value of property characteristic

σ_x standard deviation of property characteristic

Response surfaces created based on *zscores* that are both negative and positive will not predict realistic values, so all *zscores* have to be transformed into positive values by adding an integer smaller than the minimum *zscore* (i.e. if negative). The resultant *zscores* are then divided by the integer so as to maintain scale. *Zscore_char* is created using property characteristics that are highly correlated with location (*x,y*). Each characteristic will have its own *zscore_char,* so to combine all *zscore_chars* a mean *zscore_char* is computed.

The two *zscores,* geographic coordinates and property identification number (PID), are exported to *ArcGIS* for visualisation and analysis purposes. Once displayed in *ArcMap* the *zscores* are classified into nine intervals using the manual method. The *zscores* are then explored for normality and trends that could be imbedded in the data using the *geostatistical extension module*. The *histogram tool* is used to check that both *zscores* follow a normal distribution, and where they do not, a transformation is applied. The *normal qq plot* is used to confirm that the data lies on a straight line if it follows a normal distribution and the *trend analysis tool* is used for checking trends within the data.

After the checks have been carried out the next step is to analyse the effects of location using the four approaches listed below:

- *Traditional Approach* – delineated neighbourhoods.
- *Geographic Stratification* – based on a physical feature that is Table Mountain.
- *Clustering based on mean SP*– homogenous neighbourhoods.
- *Single Model* – response surfaces.

4.2.2.1 Location based on Delineated Neighbourhood Values

The neighbourhood with the median SPSM is chosen to be the average neighbourhood. Thornton with a median value of 1997 is chosen and the corresponding scalar values for all the neighbourhoods are computed using equation 4.2. Each property is given a location factor equivalent to the scalar value of its neighbourhood and the base models incorporated with these location factors are recalibrated.

4.2.2.2 Geographic Stratification

Table Mountain is used as the geographic feature for analysing locational effects. Properties surrounding Table Mountain are placed in a group known as around Table Mountain (ATM), whereas those not around it form another group known as not around Table Mountain (NATM). In this approach of accounting for location using geographic stratification, the base models ATM and NATM are treated as having locational effects implicitly incorporated in them.

4.2.2.3 Homogenous Neighbourhoods

The neighbourhoods are split into six homogenous groups, based on the mean sale price as shown in table 4.5. Different hybrid models are then specified and calibrated based on these six groups.

Table 4.5 Homogenous neighbourhoods

Neighbourhood	Mean SP Interval	Group
Hanover Park, Tafelsig, Heideveld, Parktown, Factretown, Beacon Valley, Lentegeur, Eastridge, Rocklands, Kewtown, Bridgetown, Woodlands, Silvertown	Up to R100K	1
Portland, Westridge, Weltevreden Valley, Strandfontein West, Woodstock Low, Mandalay, Strandfontein East, Primrose Park, Salt River Low, Newfields, Maitland, Kensington, Athlone, Hazendal, Welcome, Pinati, Vanguard, Belgravia, Ndabeni, Surrey, Belthorn Estate	+R100K to R200K	2
Mountview, Lansdowne Low, Woodstock Middle, Gatesville, Penlyn Estate, Lansdowne High, Thornton, Salt River High, Schotschekloof, Crawford, Observatory, Rosebank Middle, Sybrand Park, Rylands, Rondebosch East, Zonnebloem	+R200K to R300K	3
Mowbray, Woodstock High, Kenilworth Low, Claremont Low, Mouille Point, Cape Town City Centre, Pinelands, Rondebosch Low	+R300K to R500K	4
Three Anchor Bay, Gardens Lower, Vredehoek, Oranjezicht Lower, Rosebank Upper 2, Green Point, Rosebank Upper 1, Sea Point, Tamboerskloof, Claremont High, Rondebosch High, Newlands, Kenilworth High	+R500K to R1M	5
Oranjezicht Upper, Gardens Upper, Fresnaye, Camps Bay Upper, Bantry Bay, Camps Bay Bakoven, Clifton, Bishopscourt	+R1M	6

4.2.2.4 Response Surfaces

The datasets shown in table 4.6 are imported into *ArcGIS*. The *modelling sample* is in point form whilst the *properties* file is a polygon file, so a new polygon file encompassing the characteristics for the *modelling sample* is developed through the *selection by location* procedure in *ArcGIS*. This new file will act as a storage medium for the predicted location factors and is named *sold Cape Town properties*. The neighbourhoods for the Cape Town MLC are chosen through the *selection by attribute* procedure from the *neighbourhoods* file and are given the name *Cape Town neighbourhoods*.

TABLE 4.6 Datasets

File	Remarks
Modelling sample	Data to be used in the modelling
Test sample	Data to be for validation
Neighbourhoods	Shapefile for Unicity neighbourhoods
Properties	Shapefile for Unicity properties

Two 3D surfaces showing the areas of high *(peaks)* and low *(dips)* value are created for both *zscores* using the *Spatial Analysis* tool. The procedure is as follows:

- Add *Spatial Analysis* extension on *ArcMap*
- Use convert features *(zscores)* to raster *(image)* option
- Convert property/neighbourhood boundary shapefile to dxf using *ArcTools*
- Open *3D Scene*
- Add image and boundaries
- Go to properties dialogue box of both files and change base heights from constant to those of image resulting in a 3D surface.

Ten surfaces using different spatial interpolation techniques are then created and they are:

Value Influence Centres (VIC) using NCSS

A contour map of the modelling sample is created in *NCSS* using the *contour plot module*. Locations with the highest and lowest values are identified and are as shown in table 4.7. The coordinates have to be scaled down because of their large magnitudes using the transformations:

$$X\_Trans = \frac{X}{10000} \text{ and } Y\_Trans = \frac{Y}{1000000}$$

The next step off generating location factors is achieved in the *response surface analysis module*. The reference point for the distance variable is set to both the highest and lowest point, table 4.7.

The distance from the reference point to each sold property is calculated using equations 4.5 and 4.6, and together with the transformed coordinates a location factor can be estimated for each record. The location factors are then incorporated into the base models and recalibrated.

$$Highest-dist = \sqrt{\left[(X\_Trans-(-57436.00))^2 + (Y\_Trans-(-3756331.24))^2\right]} \quad 4.5$$

$$Lowest-dist = \sqrt{\left[(X\_Trans-(-34549.65))^2 + (Y\_Trans-(-3770789.97))^2\right]} \quad 4.6$$

TABLE 4.7 Value Influence Centres

Point	X	Y
Highest	-57436.00	–3756331.24
Lowest	-34549.65	–3770789.97

Trend Surface Analysis (TSA)

A non-linear regression model is developed using the *non-linear regression module* in *NCSS*. It consists of the transformed coordinates developed for the value influence centre model being added to the property characteristics, thereby resulting in a polynomial. The transformed coordinates are continuously raised to a higher power until there is no change to the predicting accuracy of the model.

Inverse Distance Weighting (IDW)

The IDW interpolation technique tool resident in the *geostatistical extension module* is used to create IDW surfaces for both *zscores*. The power of the weighting function can be specified manually or automatically (i.e. the IDW algorithm searches for the optimum power) and the other parameters are set as shown in table 4.8. The response surface is generated after the parameter setting exercise has been completed. The surface will generally not cover areas without data points so it is stretched to cover the rectangular extent of the Cape Town neighbourhoods and this can be achieved by using the *extent function*. The surface is then clipped so that it follows the outline of neighbourhoods using the *clip tool* under the *frame properties module*. The next step is to estimate location factors and the process is as stated below:

- Right click on IDW prediction map
- Click prediction
- Specify input file as *Sold Cape Town Properties*
- Specify output file and run the module

- Right click on the output file and open attribute table
- The values under the predicted field are the location factors

The location factors are exported from *ArcGIS* back to *NCSS* for recalibration with the base models and the process is as stated below:

- Click options on the attribute table
- Select export
- Specify output file
- Import file into *NCSS*
- Assign location factors to specific properties using *LOOKUP* command
- Re-run or recalibrate the model with the location factor

Local Polynomial Interpolator

The power of the interpolator is set to 1 as it is found to produce predictions that have smaller errors compared to higher powers, where the errors tend to increase. The steps used in creating an IDW surface are also applied in creating a local polynomial surface. The parameter settings are as shown in table 4.8.

TABLE 4.8 Parameter settings

Selected Method	IDW		Local	
	Characteristics	SPSM	Characteristics	SPSM
Method Parameter(s)				
Power	1	1.0414	1	1
Weight			486.84655	522.56729
Searching Neighborhood				
Neighbors to Include	4 (at least 2)	4 (at least 2)	4 (at least 2)	4 (at least 2)
Maximum	12			
Searching Ellipse				
Angle of major axis	305	305	305	305
Major Semi axis	100	100	100	100
Minor Semi axis	50	50	50	50
Sector Mode	1	1	1	1

Global Polynomial Interpolator

The power of the polynomial is set to 6 since it produces smaller prediction errors compared to lower orders. Higher orders give rise to an unstable surface, which fails to predict estimates. The same steps used to develop an IDW surface are also employed in generating the global polynomial surfaces for both the property characteristics and SPSM *zscores*.

Ordinary Kriging Interpolator

The first step is to search for the greatest distance between the properties that have sold within each neighbourhood so that a lag distance can be determined. The largest distance is found to be 700 metres, so the lag distance is set to 50 metres resulting in 14 lags (i.e. 700m÷50m). The lag distance is chosen based on the extent of the properties that is an average length of 45 metres and a width of about 30 metres. The orientation of the major axis of the neighbourhood ellipse is set to 305 degrees measured from north in a clockwise direction. The angle is chosen because the property prices increase from the south eastern side to the north western side, that is towards the mountain. By orienting the major axis in this direction it is hoped the location factor will be able to capture the general sale price trend.

The semi major axis distance is set to 100 metres because it is not expected that properties over 100 metres either side of the property, whose market value is being estimated, will have any influence on its final value, as the property sizes range from 12 metres by 20 metres in Mitchell's Plain to 50 metres by 70 metres in Bishopscourt.

The *Log transformation* is applied to the *zscore_ SPSM,* so as to transform their distribution, from a positive skewed distribution to a normal distribution, whereas the *zscore_char is* left untransformed since it already follows a normal distribution. The Gaussian model is used for fitting the curve through the semivariogram, because through visual inspection it provides the best fit. Mathematically the Gaussian model is a function used for normally distributed data and it follows that it should be used for fitting curves to both *zscores*, which follow a normal distribution. The ordinary kriging parameter setting for both *zscores* can be seen in *Appendix I.* The same steps applied in creating the IDW surface are also applied in the ordinary kriging process

4.3 Test file

A new polygon file for the test sample known as *test file* is created in a similar manner to the *Cape Town sold properties* file so as to enable location factors to be assigned to it. The linearised values derived from the modelling sample are applied to the test sample using the *LOOKUP* command in *NCSS*. All processes conducted on the modelling sample and parameters derived from the modelling sample are applied to the test sample without alteration so as to enable checks for validity (i.e. whether the same results are obtained for models) and reliability to be carried out.

CHAPTER FIVE

Results and Analysis

This chapter focuses on the results obtained from the exploratory data analysis phase through to the analysis of location. The exploratory phase looks mainly at the computation of *zscores* and the distribution (i.e. either normal or skewed) they follow prior to being used for creating response surface. The modelling section is concerned with the different models developed, which are: Cape Town (CT), around Table Mountain (ATM) and not around Table Mountain (NATM). Also, the results produced by the models incorporated with location factors generated from different techniques are examined. The models can be assumed to be capable of producing high quality results if the appraisal ratios meet the set International Association of Assessing Officers (IAAO) standards, that is the coefficient of dispersion (CoD) and coefficient of variation (CoV) values must be below 15%, and the range of the price related differential (PRD) should lie between 0.98 and 1.03.

5.1 Exploratory Data Analysis

Cleaning is conducted on a neighbourhood basis because of the huge difference in the sale price (SP) of about R6,8 Million and the sale price per square metre (SPSM) of about R/m^{2}24 000 between the high value neighbourhoods and low value neighbourhoods. Properties in a neighbourhood whose SPSM values lie two or three standard deviations away from the mean SPSM of that neighbourhood are removed from the dataset. This is undertaken as in the high value areas sales tend to be more consistent than in the low value neighbourhoods. The consistency of data in the high value neighbourhoods to the low value neighbourhoods can be attributed to the fact that people in these neighbourhoods understand how the property market works and would rarely sell a property below market value resulting in a smaller number of outliers. The removal of incorrect data and outliers from the initial Cape Town MLC dataset resulted in 4120 cleaned property records remaining from a total of 6021 records.

The cleaned records are then split into a modelling file and test file based on the ratio 80:20, so that the modelling sample can have more properties (3299 records) to the test sample (821). This will ensure that the models and response surfaces are built from an adequate data sample, thereby giving rise to market value estimates that closely resemble the real situation. Each neighbourhood will contribute properties to each sample using the ratio 80:20, as this will enable representativeness in the nature and distribution of the properties within the Cape Town MLC to be maintained. The rationale for having a test sample is to enable checks for validity and reliability to be conducted on the models produced using the modelling sample.

The sales are adjusted for time using the trend sales analysis method, because to a certain extent it minimises the error due to time differences that can arise in the location factors, if they are estimated from response surfaces built using sales at different time scales (i.e. different dates). Another advantage is that it enables the inflation pattern to be visually identified as shown in figure 5.1, compared to a situation where the time adjustment is conducted during model calibration, as one cannot visually see the trend. Figure 5.1 represents the linear regression plot for the sale appraisal mean values, whereas the actual count on a monthly basis can be seen in *Appendix F*.

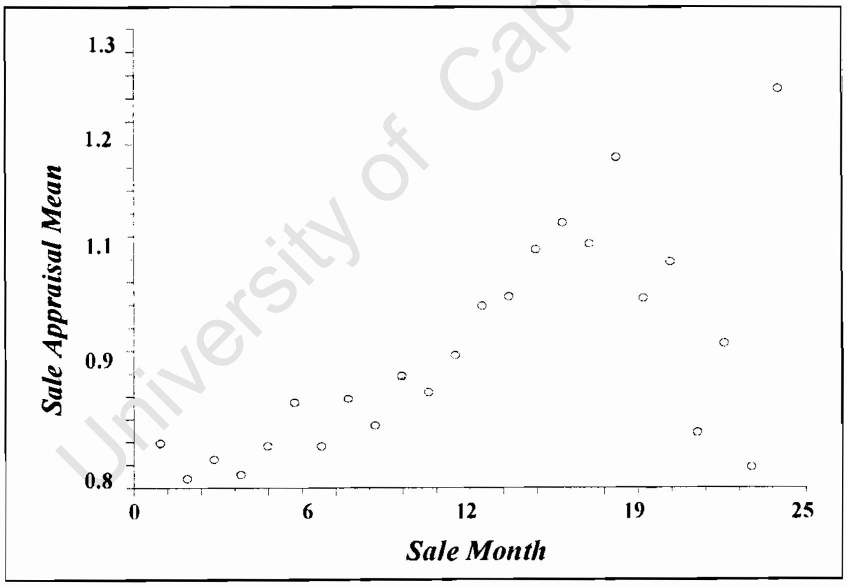


FIGURE 5.1 Plot of sale appraisal mean value on a monthly basis

The equation of the straight line is as given in equation 5.1.

$$\textit{Sale Appraisal Mean Ratio} = 0.8139 + 0.0106 \times \textit{Sale Month} \tag{5.1}$$

The increase per month, which is equivalent to the gradient, is 0.0106 and the resulting inflation on a monthly basis is represented by equation 5.2.

$$\textit{Inflation / Month} = \frac{\textit{Increase / Month}}{\textit{Intercept}} = \frac{0.0106}{0.8139} = 0.013 = 1.3\% \tag{5.2}$$

The inflation or rate is then substituted into equation 4.1, resulting in a time adjusted sale *(TAS)* function;

$$TAS = S.P[1 + r(13 - \{(YR = 1999)(Month) + (YR = 2000)(Month + 12)\})]$$

$$TAS = S.P[1 + 0.013(13 - \{(YR = 1999)(Month) + (YR = 2000)(Month + 12)\})] \qquad 5.3$$

Equation 5.3 is used to transform all the property sale prices to the assessment date. Once the sale prices are transformed, the qualitative property characteristics are then linearised and the corresponding scalar values can be seen in *Appendix C*.

5.1.1 Computation of *zscores*

Based on Sale Price per Total Living Area (SPSM)

The mean and standard deviation of the SPSM are given as R/m^{2}2459 and R/m^{2}1635 respectively, by substituting these values into equation 4.3 gives rise to equation 5.4, which enables each properties zscore to be derived from the SPSM (*zscore SPSM).* The standard deviation is large because it is derived from all properties in the Cape Town MLC whose SPSM values vary from as low as R/m^{2}500 to over R/m^{2}20 000 and in this situation it cannot be used as a measure of quality of the data.

$$zscore\ SPSM = \frac{SPSM - 2459}{1635} \qquad 5.4$$

The minimum *zscore SPSM* is found to be -1.149, whereas the maximum is 14.002, but a value of two is added to all *zscore SPSM* so that they can be transformed to positive values and divided by two to maintain scale. This results in new minimum and maximum values of 0.425 and 8.001 respectively.

Based on Property Characteristics

The correlation matrix shown in table 5.1 shows the relationship between location (*x,y*) and the property variables. Total living area (*TLA),* total fixtures *(totfix)* and garden (*gard)* are highly correlated with the *y* coordinate as seen by their high correlation coefficients of -0.605 *(r^2=37%)*, -0.532*(r^2=28%)* and –0.500*(r^2=25%)*.

TABLE 5.1 Correlation matrix

	X	Y
X	1	-0.80722
Y	-0.80722	1
STR	-0.10541	0.077965
PARK	0.283352	-0.29437
SEC	-0.44507	0.348153
TOPO	-0.3396	0.21516
VIEW	-0.45729	0.25475
GARD	-0.49954	0.437078
DTYPE	-0.26886	0.160517
CWALL	-0.12321	0.073443
COND	-0.25162	0.145014
QUAL	-0.39099	0.242226
POOL	-0.24917	0.213215
TLA	-0.60526	0.467408
TOTFIX	-0.53161	0.434914
GFSQ	-0.13943	0.132944
GARCP	-0.3138	0.27663
STOREYS	-0.14547	0.066262
ROOFING	-0.13936	0.046952
EXWALLS	-0.42986	0.337896
ERFEXT	-0.32493	0.29807

Variables such as street *(str)*, number of storeys in a dwelling *(storeys)* and type of roof cover *(roofing)* are poorly correlated with the x coordinate as seen by their low correlation coefficients of 0.078$(r^2=0.5\%)$, 0.066$(r^2=0.4\%)$ and 0.047$(r^2=0.2\%)$ respectively. The variables total living area *(TLA)*, total fixtures *(totfix)*, garden *(gard)*, view *(view)*, quality of materials used in constructing the dwelling *(qual)*, external walls of dwelling *(exwalls) and* security *(sec)* are used to compute each properties zscore *(zscore char)* using equation 4.4. The property variables used in computing *zscore char* are highly correlated with location, since most dwellings in the high value areas have high quality characteristics and vice versa for the low value areas. The corresponding standard deviations, means, minimum and maximum values of the variables are as shown in table 5.2. All the minimum values of the *zscore char* are negative and higher than -2, so a value of two is added to all of them and the resultant value divided by the integer two. The transformed seven *zscore char* are added together to get a mean *zscore char* with minimum and maximum values of 0.448 and 2.653 respectively.

TABLE 5.2 Statistics for property variables

	Mean	Deviation	Lowest	Highest
SEC	1.122293	0.236627	-0.7434	2.6571
VIEW	1.084855	0.247075	-0.8291	4.5965
GARD	0.803587	0.18109	-0.8213	3.5655
QUAL	1.012089	0.210012	-1.4648	4.4456
TLA (m^2)	122.4838	65.73461	-1.27	7.2643
TOTFIX	4.438314	3.089439	-1.4366	6.3318
EXWALLS	0.944294	0.154309	-1.83	9.55

5.1.2 Distribution of *zscores*

Figure 5.2 shows the distribution of the *zscores* derived from the SPSM and it can be seen that they follow a positively skewed distribution. The positively skewed distribution can be attributed to the fact that the majority of the properties in the dataset come from the low value neighbourhoods. Figure 5.2 helps in verifying the statement above as it can be seen that about 80% of the properties fall within the 0.43 to 1.19 interval, whereas the maximum *zscore SPSM* is eight. The high variation in the values of the *zscore SPSM* can be attributed to properties situated along Table Mountain and the Atlantic Seaboard, which have extremely high values compared to properties in other parts of the Cape Town MLC. In order for the data to be used in the kriging interpolator it has to be transformed to follow a normal distribution, and this can be achieved using the *Log transformation method,* whereas use of the *Box-Cox transformation* will result in a skewed distribution. The normally distributed dataset is shown in figure 5.3.

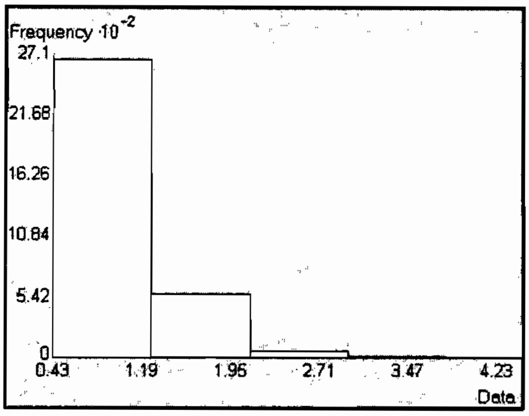


FIGURE 5.2 Distribution of SPSM *zscores*

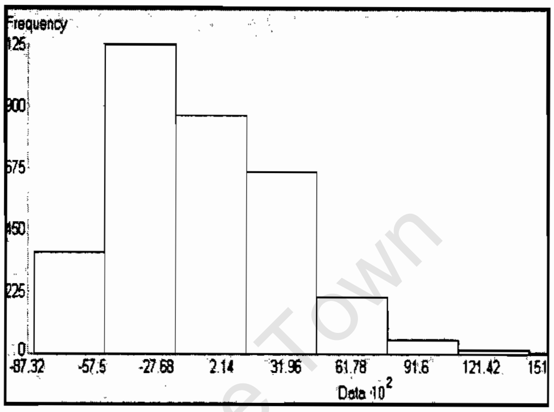


FIGURE 5.3 Transformed dataset

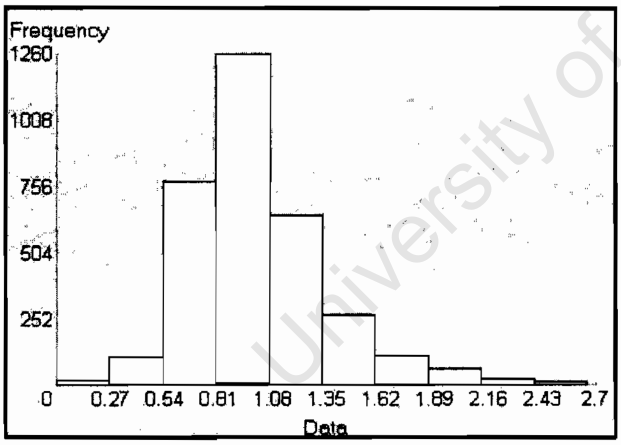


FIGURE 5.4 Property characteristics

The *zscores* derived from property variables are more consistent and follow a normal distribution compared to the *zscores* derived from the SPSM, because most of the time property characteristics will be similar within a neighbourhood, but dissimilar between neighbourhoods. No transformation is applied to the *zscore char* because the data already follows a normal distribution as seen in figure 5.4. Most of the data lies in the interval 0.54 to 1.35. It has thin tails with the value of the kurtosis being 5.02 and is almost symmetrical with a skewness value of 1.03. A normal distribution will generally have a skewness value of zero and a kurtosis value of three.

5.1.3 Trend Analysis

Figure 5.5 illustrates the trend that exists for property prices. Prices are increasing from the south-eastern corner towards the north-western area of the Cape Town MLC, that is they increase with a decrease in the x coordinate. The eastern half of the *xy plane* shows a steady increase in sale price, whereas the western half shows a lot of variation. The trend for property characteristics is illustrated in figure 5.6, and is similar to that of property prices. The curves on the *xz* and *yz plane* in both figure 5.5 and figure 5.6 follow a second order polynomial therefore mathematically defining the general trend of the sales data.

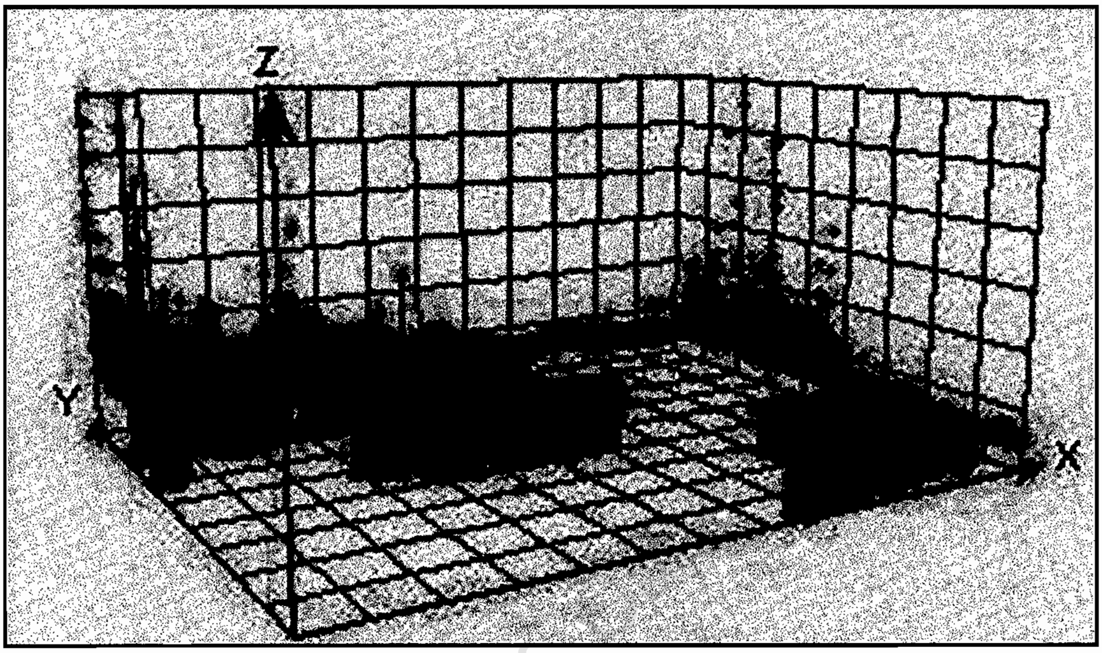


FIGURE 5.5 Trend for sales data

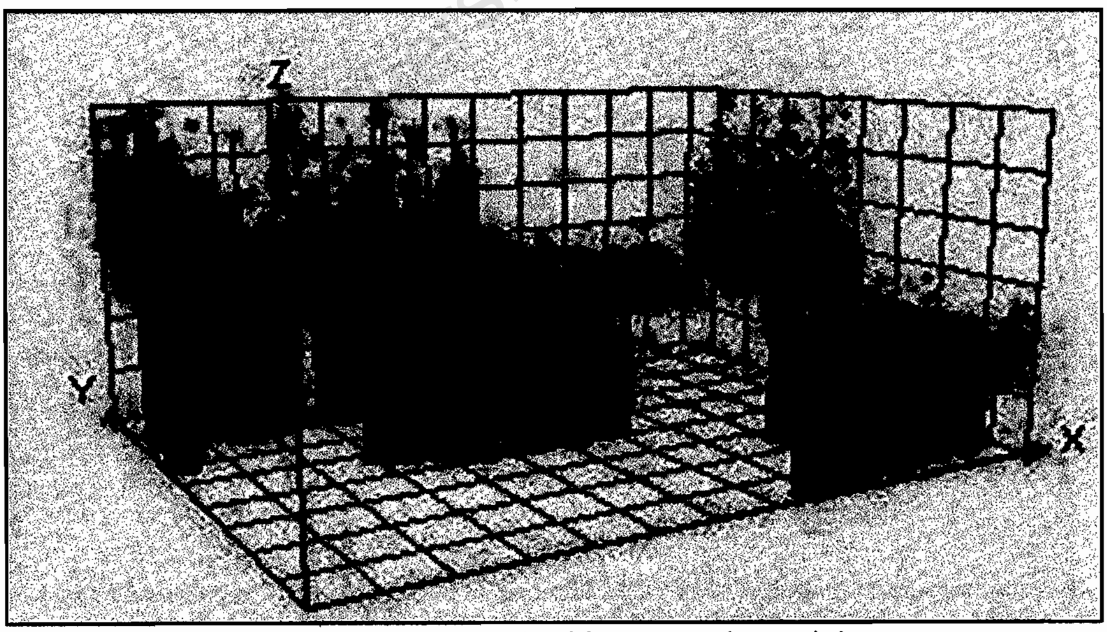


FIGURE 5.6 Trend for property characteristics

5.2 Modelling

5.2.1 Correlation Matrix

TABLE 5.3 Correlations

	SALE
SALE	1
STR	0.11
PARK	0.02
SEC	0.48
TOPO	0.39
VIEW	0.57
GARD	0.56
DTYPE	0.3
CWALL	0.28
COND	0.46
QUAL	0.54
POOL	0.4
TLA	0.79
TOTFIX	0.64
GFSQ	0.15
GARCP	0.44
STOREYS	0.32
ROOFING	0.29
EXWALLS	0.31
ERFEXT	0.63

Table 5.3 shows the relationship between the property variables and the sale price. The sale price is highly correlated with the total living area (*TLA)*, total fixtures *(totfix)* and extent of the property *(erfext)* with correlation coefficients of 0.79 *(r^2=0.62)*, 0.64 *(r^2=0.41)* and 0.63 *(r^2=0.62)* respectively. It is also highly correlated with view *(view)*, garden (*gard)* and quality *(qual)*. The correlation between the sale price and parking *(park)* is very low and this also holds for street *(str)*. Although the variables are highly correlated with the sale price, they are also correlated with each other, for example *quality* and *condition* have a value of 0.58 *(r^2=0.34)* and since they cannot be used in the model together, a new variable known as *cond_ qual* is created by multiplying the *quality* and *condition* values. The other variables that are highly correlated with each other are *total fixtures* and *garden, total living area* and *garden* with coefficients of 0.52*(r^2=0.27)* and 0.62*(r^2=0.38)* respectively. The full correlation matrix showing all the values can be seen in *Appendix E.*

5.2.2 Stepwise Regression

The final iteration results of the stepwise regression procedure are shown in table 5.4, whereas the full report can be seen in *Appendix G.* Backward regression is used and from table 5.4 it can be seen that omitting the two variables parking and dwelling type does not affect the amount of variance *(R^2=77%)* in the sale price explained by the other property variables. The use of either forward or backward regression makes no difference, as the modeller will still apply all variables in a trial and error process in search of a suitable model. Stepwise regression acts only as a guide to the amount of variability that should be expected when a certain set of variables are used in this case those determined by the stepwise regression procedure. The variables that really have an impact on the model can be seen to be *view, total living area, property extent* and *external walls* as seen by their percentage change values of 7.1, 6.8, 5.7 and 1.6 respectively for R^2. *Roofing type* is considered significant in model building although it has a lower percentage change of 0.0129 compared to dwelling type whose value is given as 0.0149. The reason it is included in the model is because of its higher *t*-statistic than *dwelling type*. The expected root mean square error (RMSE) is R199180.

TABLE 5.4 Backward Regression

		Standard.	R-Squared	R-Squared		Prob	Pct Change
In	**Variable**	**Coefficient**	**Increment**	**Other X's**	**T-Value**	**Level**	**Sqrt(MSE)**
Yes	**STR**	0.03	0.000874	0.026286	3.5187	0.00044	0.1744
Yes	**SEC**	0.067	0.00299	0.332996	6.5083	0	0.6321
Yes	**TOPO**	0.0718	0.00384	0.254749	7.3759	0	0.8155
Yes	**VIEW**	0.2302	0.034089	0.356869	21.9755	0	7.1346
Yes	**GARD**	0.0244	0.000302	0.491806	2.0669	0.03882	0.0502
Yes	**CWALL**	-0.0677	0.003012	0.342781	-6.5323	0	0.6369
Yes	**COND**	-0.0878	0.000392	0.949137	-2.3571	0.018476	0.0698
Yes	**COND QUAL**	0.3665	0.002242	0.983312	5.6354	0	0.4705
Yes	**QUAL**	-0.1771	0.001463	0.953364	-4.5524	0.000005	0.302
Yes	**POOL**	-0.0188	0.000244	0.307654	-1.8602	0.062953	0.0377
Yes	**TLA**	0.3602	0.032582	0.748877	21.4841	0	6.8286
Yes	**TOTFIX**	0.0877	0.003472	0.548896	7.0129	0	0.736
Yes	**GFSQ**	-0.0229	0.000494	0.06116	-2.6451	0.008206	0.0919
Yes	**GARCP**	-0.0489	0.001433	0.400551	-4.505	0.000007	0.2954
Yes	**STOREYS**	0.0335	0.000815	0.272645	3.3979	0.000687	0.1616
Yes	**ROOFING**	0.0123	0.00013	0.144738	1.3579	0.174601	0.0129
Yes	**EXWALLS**	0.0959	0.007389	0.196682	10.2314	0	1.5773
Yes	**ERFEXT**	0.2529	0.027108	0.57623	19.5964	0	5.7097
No	**PARK**		0.000045	0.226511	0.7996	0.423978	0.0055
No	**DTYPE**		0.000002	0.2699	0.1678	0.866761	0.0149
R-Squared = 0.769878 **Sqrt (MSE)** = 199180.2							

5.2.3 Base Models

The variable estimates for the Cape Town (CT), Around Table Mountain (ATM) and Not Around Table Mountain (NATM) base models are shown in Table 5.5. The reason for having three base models is to enable the researcher to quantify the merits of stratifying based on geographic location and homogenous neighbourhoods or just to apply a single model to the whole jurisdiction. The most significant variable within all the three models is *total living area*, which has a minimum *t-statistic* of 13 for the ATM model, whereas CT and NATM have 29 and 26 respectively. The other variables that are also significant and found within all the three models are *view* used as an overall rate multiplier (i.e. either increases or decreases the estimate), *external walls, total fixtures* and *property extent*. *Security* is used as a qualitative land variable in all models compared to street, which is more significant when used as an overall variable in the ATM model. The variables *garden, pool* and *garage* are only used in the NATM model because in the other models they are found to be insignificant, whereas *dwelling storeys* and *roofing* have an impact when dealing with the CT properties. In the ATM model, *parking, dwelling type* and *common wall* do not contribute to the outcome of the market value of a property compared to their effect in the CT and NATM models where their *t-statistics* values are greater then ±2. All base models are built using the hybrid model method because in a complex property market like Cape Town, where property characteristics and prices vary over short distances, it is the only model that can really explain such effects.

The estimated market value in the CT and NATM base models is explained more by the property variables compared to the market value estimates produced by the ATM model. This can be attributed to the fact that in the ATM region atypical property characteristics will tend to be the norm for example roof coverings can range from tiles to thatch within the same street, whereas in the NATM region property characteristics within a neighbourhood will tend to be similar or typical. The assessment ratios of the three models are shown in *Appendix J.* The CT and NATM models both have a R^2 of about 78% compared to the ATM model with a percentage of 64. The price related differential (PRD) for CT model is very high at 120% compared to the other two models with the ATM having a low of 107%, implying that all base model are regressive. The coefficient of dispersion (CoD) and coefficient of variation (CoV) are high in all three models with minimum values of 22% and 30% respectively. The models produce similar values for both the predicted and actual sales with regard to the median and mean sales. Predicted minimum sales prices are higher than the actual minimum sales and vice versa for the maximum sales, this is due to the regressivity of the models that is undervaluing of high value properties with respect to low value properties.

TABLE 5.5 Base models

		CCT		ATM		NATM	
Variable Type	**Variable Name**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**
Overall	**Park**	-2.778	-6.059			-1.366	-3.410
	View	0.621	22.338	0.438	7.971	0.498	14.406
	Str			0.449	2.214		
Qual Blding	**Cond Qual**	0.258	12.404	0.174	4.648	0.107	3.140
	Dtype	0.210	4.795			0.593	8.570
	Cwall	0.762	5.372			-0.297	-2.656
	Storeys	0.194	4.024				
	Roofing	0.184	3.574				
	Exwalls	0.952	23.674	0.805	9.893	0.707	8.860
Quan Blding	**Tla**	1747.878	28.518	2370.730	13.214	1200.367	25.853
	Totfix	17036.610	13.311	23727.520	5.929	8643.204	11.786
Qual Land	**Str**					1.728	4.768
	Topo	1.460	8.693	0.897	2.627		
	Sec	2.709	10.228	1.385	3.654	1.501	12.049
Quan Land	**Erfext**	47.994	6.030	114.986	3.975	119.508	12.007
Qual Other	**Gard**					1.650	3.742
Quan Other	**Pool**			-1800.860	-2.012	1063.302	6.267
	Garcp					627.532	4.876

5.2.4 Location analysis

Location is analysed using four approaches stated below, as they will enable conclusive results to be obtained on which technique should be applied in the Cape Town MLC for generating location factors.

- Delineated neighbourhoods
- Geographic location
- Homogenous neighbourhoods created based on mean sale price intervals
- Single model with location adjustments (i.e. response surfaces)

5.2.4.1 Delineated Neighbourhoods

Delineated neighbourhoods are investigated because they where the most common approach for analyzing locational effects, prior to the advent of spatial analysis techniques. Table 5.6 shows the results of using neighbourhood (NBHD) scalar values as location factors. The location factor is significant in all three models as seen by the high t values in the modelling and test samples. Variability of the estimated market values is well explained by the independent variables as seen by the high R^2 values that are above 75%. The NATM model is consistent for both the test sample and modelling sample with R^2 values of about 90%. The assessment ratio statistics also confirm that the NATM model is the most consistent as evidenced by the values of the CoD, CoV and PRD, which are 15.6%, 20% and about 101% respectively for both the modelling and test samples.

The CoD and CoV values of the ATM, and CT models are increasing from the modelling to the test sample. The CT model is progressive and the NATM model is regressive for both samples, whereas the ATM model is regressive and progressive with regard to the modelling and test sample. Progressivity means that when the model estimates values of highly priced properties, the estimates will be higher than the actual sale. The results produced by the delineated neighbourhoods approach do not meet the set IAAO standards. The reason for this is that a single location factor is assigned to properties within a particular neighbourhood, thereby failing to capture microvariations that occur within the neighbourhood, for example proximity to a shopping mall.

TABLE 5.6 NBHD location factor

	CCT		ATM		NATM	
	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value	
Name	(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)	
NBHD	57.7	26.6	20.8	7.1	59.6	26.4
R-Squared	0.89	0.95	0.77	0.89	0.91	0.90
	Ratio		Ratio		Ratio	
No of Cases	3279	812	611	149	2668	663
Mean	0.96	0.96	1.01	0.93	0.98	1.00
Median	0.93	0.94	0.98	0.90	0.95	0.98
CoV	0.23	0.27	0.27	0.30	0.20	0.20
CoD	18.07	21.06	20.73	24.53	15.62	15.55
Weighted Mean	0.97	0.96	0.97	0.96	0.98	0.98
PRD	99.2	99.5	103.8	97.1	100.7	102.0

5.2.4.2 Geographic Stratification

The results for geographic stratification are replicas of those discussed in section 5.2.3 for base models. Geographic stratification is conducted so as to assess the impact a single model (CT) has on the predictive accuracy compared to the two models (ATM and NATM) developed for slightly homogenous areas using Table Mountain as the geographic feature.

5.2.4.3 Homogenous Neighbourhoods

The six homogenous models are created based on the average sale price because, this is assumed to be the best approach since no prior knowledge of the extent or quality of the property characteristics within each neighbourhoods existed. The mean sale price intervals defined in chapter four are used to develop the six homogenous groups and the results produced by each model are shown in table 5.7. Variability of the estimated market values is poorly explained by the non-location variables in group one, group two and group six models as seen by their low R^2' values of below 50% compared to group four and group five models, which have values above 70%.

Generally, the CoD and CoV values are above 15% and 20% respectively for all the six groups with the exception being group four with a low CoD and CoV of 14% and 18% respectively. The high statistics can be attributed to linearised variables such as *street*, which have a reverse effect in some neighbourhoods. For example, properties close to a busy street or expressway can be more valuable in a poorer neighbourhood because of easy access to public transport. All group models are regressive and have an average value of 102%. The assessment ratios for the mean and median are close to one. The most consistent model in this section is that for group four.

TABLE 5.7 Homogenous NBHD

	Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
	Model	Test	Model	Test	Model	Test	Model	Test	Model	Test	Model	Test
R-Squared	0.38	0.55	0.65	0.72	0.46	0.47	0.73	0.7	0.7	0.73	0.46	0.79
	Ratio		Ratio		Ratio		Ratio		Ratio		Ratio	Ratio
No of Cases	598	149	815	200	595	148	606	151	494	123	176	41
Mean	1.02	1.03	1	1.02	1.01	1.02	1	1.01	1.01	0.97	1.13	1.11
Median	0.98	0.99	0.97	0.98	0.99	1	0.97	1	0.98	0.94	1.08	1.03
CoV	0.21	0.2	0.2	0.22	0.21	0.2	0.18	0.18	0.22	0.23	0.38	0.51
CoD	15.51	15.57	15.84	16.77	16.57	15.09	14.23	14.12	17.78	19.06	27.55	42.84
Weighted Mean	0.99	1	0.98	0.99	0.98	0.99	0.98	0.99	0.98	0.96	1.00	1.00
PRD	102.9	103.1	101.9	103.1	102.7	103.2	101.3	102	103	100.6	112.7	111.4

5.2.4.4 Location Value Response Surfaces (LVRS)

Value Influence Centres (VICs)

The results for using the distance variable to determine location factors are shown in table 5.8. The distance to the point of lowest property value is used as the distance variable, because even if the distance to the point of highest value is applied it makes no difference as the same results are produced. Location is significant in all three models, and the estimated market values are highly explained by the independent variables as seen by the high R^2 values. The models produced are regressive save for the test sample for the ATM model. Analysing location using the VIC technique results in models that produce assessment ratios (CoD and CoV) that do not meet the set IAAO standards. This method probably could work better if the Cape Town MLC had a single economic centre (i.e. the Cape Town central business district) and not several economic sub-centres, such as shopping malls. To overcome this problem, the best approach would be to identify all centres of influence within the MLC and assign each one its own distance variable to properties within that area. Logically, this approach makes sense, but practically it is impossible because of limited resources such as time and the need to minimise the complexity of the model.

TABLE 5.8 VIC results

	CCT		ATM		NATM	
	Model	**Test**	**Model**	**Test**	**Model**	**Test**
Variable	**T-Value**	**T-Value**	**T-Value**	**T-Value**	**T-Value**	**T-Value**
Name	**(Ho: Bi=0)**	**(Ho: Bi=0)**	**(Ho: Bi=0)**	**(Ho: Bi=0)**	**(Ho: Bi=0)**	**(Ho: Bi=0)**
LocMult	30.89	24.53	7.49	2.94	34.76	18.74
R-Squared	0.83	0.94	0.67	0.86	0.85	0.86
	Ratio	**Ratio**	**Ratio**	**Ratio**	**Ratio**	**Ratio**
No of Cases	3279	812	611	149	2668	663
Mean	1.11	1.05	1.04	0.96	1.03	1.03
Median	1.05	1.00	1.00	0.90	0.97	0.99
CoV	0.29	0.32	0.27	0.30	0.27	0.28
CoD	23.24	24.69	20.87	25.84	21.44	21.33
Weighted Mean	1.01	0.99	0.99	0.97	0.99	1.00
PRD	109.5	106.3	105.3	99.0	103.5	103.6

Trend Surface Analysis (TSA)

The variables highly correlated with the sale price *(view, qual, TLA, totfix, gard and erfext)* are used to develop a model to which a polynomial consisting of the transformed coordinates *(x trans,y trans)* is added. The transformed coordinates are preferred to the original coordinates because they are smaller in terms of magnitude, thereby allowing for faster processing times and also enabling the polynomial to be raised to higher orders. A non linear multiple regression model is developed for the whole Cape Town MLC and is represented by equation 5.5. The polynomial is of the third order and consists of four variables that is x, y, x^2 and x^3. The x coordinate is used as the major variable in the polynomial, since it is raised to higher powers. The reason for this is that a greater variation in property prices occurs in an east-west direction, this corresponds to the x-axis. The model produces an R^2 value of 81% for the modeling sample and 91% for the test sample. Although TSA produces very high R^2 values, its major weakness is the inability to generate other statistical measures such as CoV and CoD to enable an evaluation of model accuracy. It is also very unstable in that if the polynomial is raised to the fourth order it begins to generate errors, and the time taken conducting computations is well over ten minutes for a third order polynomial.

$$MV' = -40737180 + view^{0.484}\begin{bmatrix} qual^{0.501}(1559 * TLA + 14793 * Totfix) \\ + gard^{0.692}(245 * Erfext) \end{bmatrix} + \begin{bmatrix} -3350410x - 9631589y \\ -868365x^2 - 75895x^3 \end{bmatrix} \qquad 5.5$$

3D Response Surface of zscore SPSM

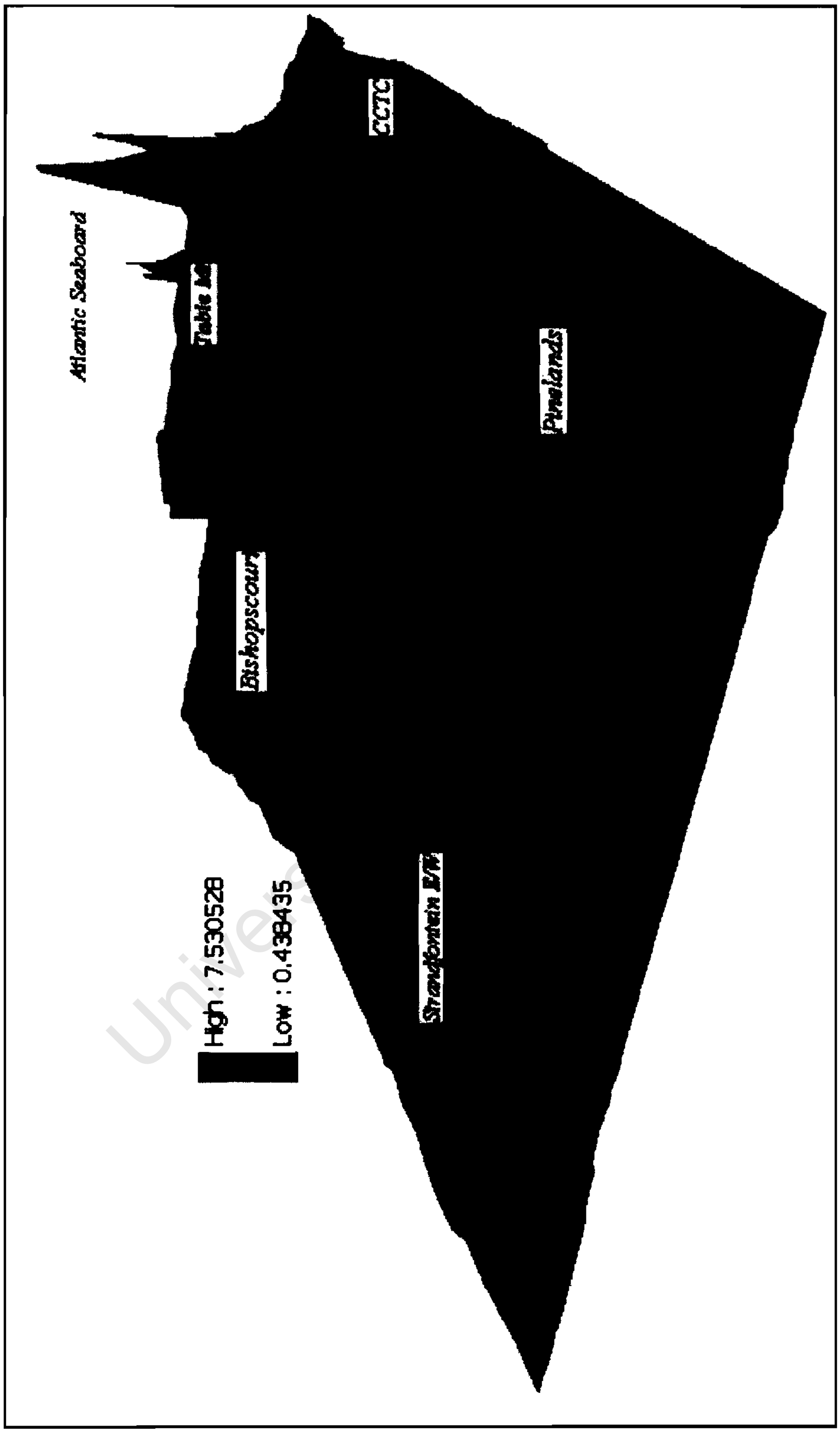


FIGURE 5.7 3D LVRS for *zscore SPSM*

Figure 5.7 represents a 3D surface for the *zscore* derived form the sale price per square metre. There is a huge difference in value between the neighbourhoods in the lower left corner and those on the upper right hand corner representing properties facing the Atlantic Ocean. The properties of high value (*i.e. properties in the red zone)* are those surrounding Table Mountain, especially in neighbourhoods like Clifton, Camps Bay and Bishopscourt.

The results using different techniques for generating location factors in the CT model are shown in table 5.9. All the location factors derived from the different techniques are significant as seen by their high *t* values except for those generated from the inverse distance weighting (IDW) surface, where the modelling sample produces a value of zero. At least 85% of the variability in the estimated market values of the properties is explained by all variables, with all location factors generated from the different techniques accounting for most of the variability. CoV and CoD values are generally above 20% except for ordinary kriging with values of about 15% and 17%.

All techniques result in the final model being regressive because they all have PRD values above 105%, except for ordinary kriging, which produces a PRD value of about 100%. The mean and median ratios are also close to one. The surfaces created using the global interpolator and ordinary kriging technique produce results that are consistent between samples, although those of ordinary kriging are much superior. The change in the CoD and CoV values between samples for the IDW method is extremely high with an increase of about 160%, whereas that for the local technique is about 60%.

The ratio statistics produced using most techniques do not meet the IAAO standards especially IDW which is very unstable. The only exception is ordinary kriging, whose ratios closely follow the IAAO specifications. A high degree of variability between ratios should be expected when modelling properties over a broad range of property markets from extremely up-market to sub-economic. Using a single model for the whole jurisdiction, even with location adjustments, will not produce desirable results especially in a very heterogeneous and complex property market in which certain variables may influence market value quite differently in different areas.

TABLE 5.9 CT Results

	IDW		Global		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value		T-Value	
Name	(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)	
LocMult	0.0	33.9	36.9	33.2	99.8	19.5	70.9	34.8
R-Squared	1.00	0.96	0.85	0.96	0.94	0.94	0.91	0.96
	Ratio		Ratio		Ratio		Ratio	
No of Cases	3268	809	3268	809	3268	809	3268	809
Mean	1.06	1.05	1.09	1.05	1.08	1.13	0.98	1.03
Median	1.04	1.02	1.04	1.02	1.05	1.10	0.96	1.01
CoV	0.09	0.23	0.27	0.27	0.18	0.29	0.20	0.22
CoD	6.89	17.32	20.92	20.43	13.36	22.45	15.42	16.87
Weighted Mean	1.01	1.00	1.01	1.00	1.01	1.01	0.98	0.99
PRD	104.9	105.1	107.8	105.7	106.4	111.8	100.4	103.4

Results for the ATM model are shown in table 5.10. The location factor is significant for all different techniques except for the IDW where it has an extremely high t value of 330. The R^2 varies a lot between the modelling and test samples of each technique. The CoV and CoD values are above 20% but those for the IDW interpolator are 3% and 1.87% for the modelling sample, but increase dramatically for the test sample to 26% and 22%. The PRD falls from above 100% to around 90% for the modelling and test samples with regard to all location techniques. Ordinary kriging produces results that have lower ratios in terms of the CoD and CoV to the other methods.

All techniques produce results that are not consistent between samples. The cause for the statistics not meeting the set IAAO standards could be due to the high variation of property prices over short distances within the ATM neighbourhoods. The variation in prices can be attributed to the variable views of either Table Mountain or the Atlantic Ocean. For example, a property without any view of either of the two features can result in it selling for a couple of hundred thousands of rands less than the ones privileged to have good views. What complicates matters further, is that the properties may be in the same street or even adjacent.

Results for the NATM model are shown in table 5.11. Every technique produces a very high R^2 of above 85%, with IDW and ordinary kriging producing values above 90%. The CoV values are generally around 20%, whereas the CoD values are good for the local polynomial interpolator and ordinary kriging technique, being 11%:16% and 13%:16% respectively with regard to the modelling and test sample. All the PRD values are above 100% but no more than 104%. All the ratios for the test and modelling samples are consistent for each technique except for IDW. The results produced using the NATM model incorporated with location factors are not of high quality and this can be attributed to the mixture of properties from the low value areas (mainly Cape Flats region) and high value areas (southern suburbs).

TABLE 5.10 ATM Results

	IDW		Global		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value		T-Value	
Name	Ho: Bi=0		Ho: Bi=0		Ho: Bi=0		Ho: Bi=0	
LocMult	330.4	11.2	10.4	10.2	35.5	6.6	26.0	12.1
R-Squared	1.00	0.92	0.69	0.91	0.88	0.89	0.82	0.93
	Ratio		Ratio		Ratio		Ratio	
No of Cases	604	147	604	147	604	147	604	147
Mean	1.00	0.96	1.04	0.94	1.02	0.95	1.01	0.96
Median	1.00	0.93	1.01	0.91	1.00	0.89	0.98	0.94
CoV	0.03	0.26	0.28	0.29	0.19	0.31	0.22	0.25
CoD	1.87	21.51	20.83	24.48	13.57	26.92	17.40	20.35
Weighted Mean	1.00	0.98	0.99	0.97	0.99	0.97	0.99	0.98
PRD	100.2	97.9	105.0	96.9	102.5	98.1	102.4	98.1

TABLE 5.11 NATM results

	IDW		Global		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value		T-Value	
Name	Ho: Bi=0		Ho: Bi=0		Ho: Bi=0		Ho: Bi=0	
LocMult	677.3	28.1	44.3	22.9	96.4	26.6	76.0	28.7
R-Squared	1.00	0.91	0.88	0.88	0.95	0.90	0.94	0.91
	Ratio		Ratio		Ratio		Ratio	
No of Cases	2664	662	2664	662	2664	662	2664	662
Mean	1.01	1.03	1.04	1.04	1.00	1.03	0.99	1.02
Median	1.00	1.00	0.99	1.00	0.98	1.01	0.97	0.99
CoV	0.04	0.20	0.24	0.25	0.15	0.21	0.17	0.20
CoD	2.37	15.53	18.87	18.13	10.91	16.10	13.05	15.54
Weighted Mean	1.00	0.99	0.99	1.00	0.99	0.99	0.98	0.99
PRD	100.9	103.4	104.1	104.0	100.8	103.9	101.0	103.2

Table 5.12 shows the results for the six homogenous groups incorporated with location factors. All groups produce high R^2 values of above 70% except for groups one and three. The models produce low CoD and CoV values of about 15% on average and PRD values of about 102%, with the mean and median statistics close to one, except for group six. Model stability is high and the results produced by both the modelling and test samples for groups one to five are consistent, and also meet the set IAAO standards. Group six performs poorly because it consists of properties from the extremely high value neighbourhoods, where values vary over short distances and this results in the model behaving in an unpredictable manner as seen by the variation of the appraisal ratios between samples. Homogenous models with location factors

produce good results because similar properties are used in each model and the location factor accounts for the microvariations within each neighbourhood.

TABLE 5.12 Homogenous neighbourhoods with location factors

	Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
	Model	Test	Model	Test	Model	Test	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value		T-Value		T-Value		T-Value	
Name	(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)	
LocMult	13.05	2.74	13.32	5.64	17.98	4.15	16.08	3.44	16.03	8.25	13.72	5.88
R-Squared	0.52	0.57	0.71	0.76	0.65	0.53	0.81	0.73	0.8	0.83	0.71	0.91
	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio
No of Cases	598	149	815	199	593	148	604	151	492	122	170	40
Mean	1.01	1.02	0.99	1.02	1	1.02	0.99	1	1	0.98	1.07	1.05
Median	0.98	0.98	0.96	0.97	0.99	1	0.97	1	0.98	0.97	1.05	1
CoV	0.18	0.19	0.18	0.22	0.16	0.18	0.15	0.17	0.18	0.18	0.27	0.35
CoD	13.14	15.36	13.7	15.95	12.59	13.79	11.59	13.34	14.04	14.53	20.41	27.95
Weighted Mean	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.98	1.01	1
PRD	102.1	102.8	101.2	102.5	101.5	102.5	100.3	101.3	101.6	100.3	106	105.1

The full report on results using location factors estimated from surfaces created using *zscores* derived from the sale price per square can be seen in *Appendix K* and the corresponding response surfaces in *Appendices M, N, O* and *P*.

Surfaces created from Property Characteristics

Response surfaces based on property characteristics are also built and this is undertaken with the intention of comparing with those based on sale prices per square metre. Figure 5.8 represents a 3D surface created for *zscores* derived from property characteristics for the Cape Town MLC. The variation is not as high as that in Figure 5.7. Generally, properties close to the mountain show a similar pattern in that they all have high values and the values decrease gradually to the lower left corner. The 2D response surfaces produced by the different interpolation techniques using the *zscores* derived from the property characteristics can be seen in *Appendices M, N, O* and *P*.

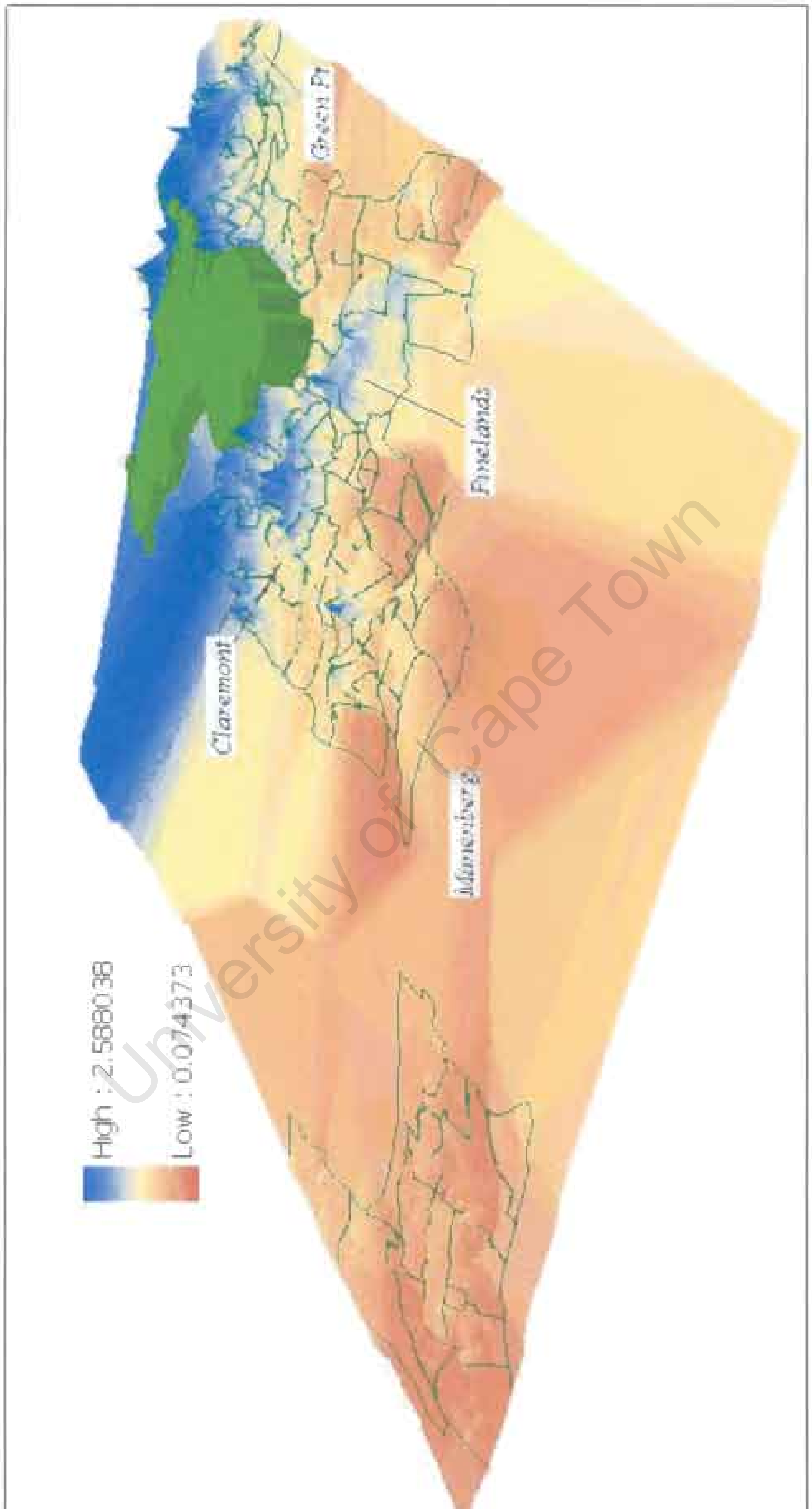


FIGURE 5.8 3D Surface of property characteristics

The t and R^2 values for the CT model are high but not consistent between modelling and test samples as seen in table 5.13. All the CoD and CoV values are high with a minimum of 23% and 30% respectively. The models produced are regressive except for those due to the global polynomial and ordinary kriging techniques, which are also progressive with regard to their test samples. The local polynomial and ordinary kriging techniques produce ratios that are similar between samples therefore making them the most consistent methods. The global polynomial and IDW interpolation techniques produce ratios that vary between samples. The mean and median ratios are close to one for the modelling samples, whereas for the test samples the ratios are around 0.9.

TABLE 5.13 CT surfaces based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	**T-Value**		**T-Value**		**T-Value**		**T-Value**	
Name	**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**	
LocMult	32.9	27.5	7.8	20.4	21.0	10.7	29.6	20.3
R-Squared	0.84	0.95	0.79	0.94	0.81	0.92	0.83	0.94
	Ratio		**Ratio**		**Ratio**		**Ratio**	
No of Cases	3268	809	3268	809	3268	809	3268	809
Mean	1.01	0.94	1.18	0.98	1.12	1.10	1.01	0.95
Median	0.97	0.92	1.10	0.94	1.05	1.05	0.96	0.92
CoV	0.30	0.34	0.36	0.32	0.33	0.33	0.32	0.32
CoD	23.31	26.11	28.90	24.97	25.94	25.90	24.74	24.93
Weighted Mean	0.99	0.97	1.02	0.96	1.00	0.99	0.98	0.96
PRD	102.5	97.3	116.2	101.2	111.4	110.7	103.3	99.0

The ATM model results are shown in table 5.14 and it can be seen that the location factor is significant when generated using the global polynomial and ordinary kriging techniques. The R^2 values are high for the test samples being about 80% compared to the modelling samples where the values are around 60%. The CoD and CoV values are generally above 20%. The model is regressive for all modelling samples but progressive for the test samples. The better technique for generating location factors is the global polynomial method.

TABLE 5.14 ATM based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Location	**T-Value**		**T-Value**		**T-Value**		**T-Value**	
Factor	**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**	
LocMult	8.2	6.0	-1.4	3.1	5.0	1.1	7.9	3.2
R-Squared	0.68	0.88	0.64	0.86	0.66	0.85	0.67	0.86
	Ratio		**Ratio**		**Ratio**		**Ratio**	
Number of Cases	604	147	604	147	604	147	604	147
Mean	1.04	0.95	1.07	0.93	1.03	0.95	1.02	0.93
Median	1.00	0.94	1.04	0.87	0.99	0.89	0.98	0.87
CoV	0.28	0.31	0.28	0.33	0.30	0.33	0.29	0.33
CoD	21.32	25.23	21.70	28.35	22.45	27.72	22.85	28.26
Weighted Mean	0.99	0.97	1.00	0.95	0.98	0.96	0.98	0.95
PRD	105.2	97.8	107.1	97.4	105.2	98.8	104.2	97.3

Results for the NATM model are shown in table 5.15. The *t-statistics* for the location factors are quite significant across the board. The R^2 values are quite high with the minimum being 79%, but the CoD and CoV values are also very high. The models are regressive with the PRD values under 105%. The global polynomial, local polynomial and ordinary kriging produce ratios that are quite consistent between samples.

TABLE 5.15 NATM based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	**T-Value**		**T-Value**		**T-Value**		**T-Value**	
Name	**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**		**(Ho: Bi=0)**	
LocMult	39.6	19.7	12.8	18.4	26.9	18.1	32.9	19.3
R-Squared	0.87	0.87	0.79	0.86	0.83	0.86	0.85	0.86
	Ratio		**Ratio**		**Ratio**		**Ratio**	
No of Cases	2664	662	2664	662	2664	662	2664	662
Mean	1.00	1.01	1.11	1.03	1.06	1.04	1.03	1.03
Median	0.94	0.97	1.06	0.99	1.01	0.99	0.98	0.99
CoV	0.27	0.28	0.30	0.27	0.28	0.27	0.27	0.26
CoD	21.21	20.78	24.33	20.58	21.92	21.12	21.06	19.95
Weighted Mean	0.98	0.98	1.00	0.99	0.99	0.99	0.98	0.99
PRD	101.8	102.2	110.4	104.4	106.7	104.7	104.9	104.2

The location factor is highly significant in all of the models and variability of the estimated market value is highly explained by the variables. The results produced using this approach do not meet the set IAAO standards, and just like surfaces created using the *zscore SPSM*, the IDW technique results in models that are unstable. The cause for poor results could be attributed to errors within the data, that is, wrong coding and missing data. Another reason could be that property characteristics are not a good indicator of what transpires in the property market compared to the actual sale price, for example if the economy is going through a recession, the property prices will sky rocket, but the property characteristics will not change. The full report on results produced using location factors estimated from response surfaces created using *zscores* derived from property characteristics can be seen in *Appendix L.*

5.2.5 Effects of Data Distribution

The distribution of data as seen in figure 5.9, has an impact on the creation of the response surface. In areas where there is no data the error is high, for example slightly above 0.16 in Strandfontein East and Strandfontein West. In areas where the distribution and size is good, such as Lansdowne low and Hanover Park, the error is smaller than 0.13. These neighbourhoods have low errors because the properties are generally homogenous with a small range of sale prices, compared to the more upmarket areas where the prices vary a lot over small distances, for example in Bishopscourt. Edge effects are also evident in the error surface, especially the lower right portion of the map where the error around the edges is above 0.17. The errors at the periphery can be attributed to interpolation techniques extrapolating values, as no sales data exists in those areas.

The t and R^2 values for the CT model are high but not consistent between modelling and test samples as seen in table 5.13. All the CoD and CoV values are high with a minimum of 23% and 30% respectively. The models produced are regressive except for those due to the global polynomial and ordinary kriging techniques, which are also progressive with regard to their test samples. The local polynomial and ordinary kriging techniques produce ratios that are similar between samples therefore making them the most consistent methods. The global polynomial and IDW interpolation techniques produce ratios that vary between samples. The mean and median ratios are close to one for the modelling samples, whereas for the test samples the ratios are around 0.9.

TABLE 5.13 CT surfaces based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable Name	T-Value (Ho: Bi=0)		T-Value (Ho: Bi=0)		T-Value (Ho: Bi=0)		T-Value (Ho: Bi=0)	
LocMult	32.9	27.5	7.8	20.4	21.0	10.7	29.6	20.3
R-Squared	0.84	0.95	0.79	0.94	0.81	0.92	0.83	0.94
	Ratio		Ratio		Ratio		Ratio	
No of Cases	3268	809	3268	809	3268	809	3268	809
Mean	1.01	0.94	1.18	0.98	1.12	1.10	1.01	0.95
Median	0.97	0.92	1.10	0.94	1.05	1.05	0.96	0.92
CoV	0.30	0.34	0.36	0.32	0.33	0.33	0.32	0.32
CoD	23.31	26.11	28.90	24.97	25.94	25.90	24.74	24.93
Weighted Mean	0.99	0.97	1.02	0.96	1.00	0.99	0.98	0.96
PRD	102.5	97.3	116.2	101.2	111.4	110.7	103.3	99.0

The ATM model results are shown in table 5.14 and it can be seen that the location factor is significant when generated using the global polynomial and ordinary kriging techniques. The R^2 values are high for the test samples being about 80% compared to the modelling samples where the values are around 60%. The CoD and CoV values are generally above 20%. The model is regressive for all modelling samples but progressive for the test samples. The better technique for generating location factors is the global polynomial method.

TABLE 5.14 ATM based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Location	T-Value		T-Value		T-Value		T-Value	
Factor	(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)	
LocMult	8.2	6.0	-1.4	3.1	5.0	1.1	7.9	3.2
R-Squared	0.68	0.88	0.64	0.86	0.66	0.85	0.67	0.86
	Ratio		Ratio		Ratio		Ratio	
Number of Cases	604	147	604	147	604	147	604	147
Mean	1.04	0.95	1.07	0.93	1.03	0.95	1.02	0.93
Median	1.00	0.94	1.04	0.87	0.99	0.89	0.98	0.87
CoV	0.28	0.31	0.28	0.33	0.30	0.33	0.29	0.33
CoD	21.32	25.23	21.70	28.35	22.45	27.72	22.85	28.26
Weighted Mean	0.99	0.97	1.00	0.95	0.98	0.96	0.98	0.95
PRD	105.2	97.8	107.1	97.4	105.2	98.8	104.2	97.3

Results for the NATM model are shown in table 5.15. The *t-statistics* for the location factors are quite significant across the board. The R^2 values are quite high with the minimum being 79%, but the CoD and CoV values are also very high. The models are regressive with the PRD values under 105%. The global polynomial, local polynomial and ordinary kriging produce ratios that are quite consistent between samples.

TABLE 5.15 NATM based on property characteristics

	Global		IDW		Local		Ordinary	
	Model	Test	Model	Test	Model	Test	Model	Test
Variable	T-Value		T-Value		T-Value		T-Value	
Name	(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)		(Ho: Bi=0)	
LocMult	39.6	19.7	12.8	18.4	26.9	18.1	32.9	19.3
R-Squared	0.87	0.87	0.79	0.86	0.83	0.86	0.85	0.86
	Ratio		Ratio		Ratio		Ratio	
No of Cases	2664	662	2664	662	2664	662	2664	662
Mean	1.00	1.01	1.11	1.03	1.06	1.04	1.03	1.03
Median	0.94	0.97	1.06	0.99	1.01	0.99	0.98	0.99
CoV	0.27	0.28	0.30	0.27	0.28	0.27	0.27	0.26
CoD	21.21	20.78	24.33	20.58	21.92	21.12	21.06	19.95
Weighted Mean	0.98	0.98	1.00	0.99	0.99	0.99	0.98	0.99
PRD	101.8	102.2	110.4	104.4	106.7	104.7	104.9	104.2

The location factor is highly significant in all of the models and variability of the estimated market value is highly explained by the variables. The results produced using this approach do not meet the set IAAO standards, and just like surfaces created using the *zscore SPSM*, the IDW technique results in models that are unstable. The cause for poor results could be attributed to errors within the data, that is, wrong coding and missing data. Another reason could be that property characteristics are not a good indicator of what transpires in the property market compared to the actual sale price, for example if the economy is going through a recession, the property prices will sky rocket, but the property characteristics will not change. The full report on results produced using location factors estimated from response surfaces created using *zscores* derived from property characteristics can be seen in *Appendix L.*

5.2.5 Effects of Data Distribution

The distribution of data as seen in figure 5.9, has an impact on the creation of the response surface. In areas where there is no data the error is high, for example slightly above 0.16 in Strandfontein East and Strandfontein West. In areas where the distribution and size is good, such as Lansdowne low and Hanover Park, the error is smaller than 0.13. These neighbourhoods have low errors because the properties are generally homogenous with a small range of sale prices, compared to the more upmarket areas where the prices vary a lot over small distances, for example in Bishopscourt. Edge effects are also evident in the error surface, especially the lower right portion of the map where the error around the edges is above 0.17. The errors at the periphery can be attributed to interpolation techniques extrapolating values, as no sales data exists in those areas.

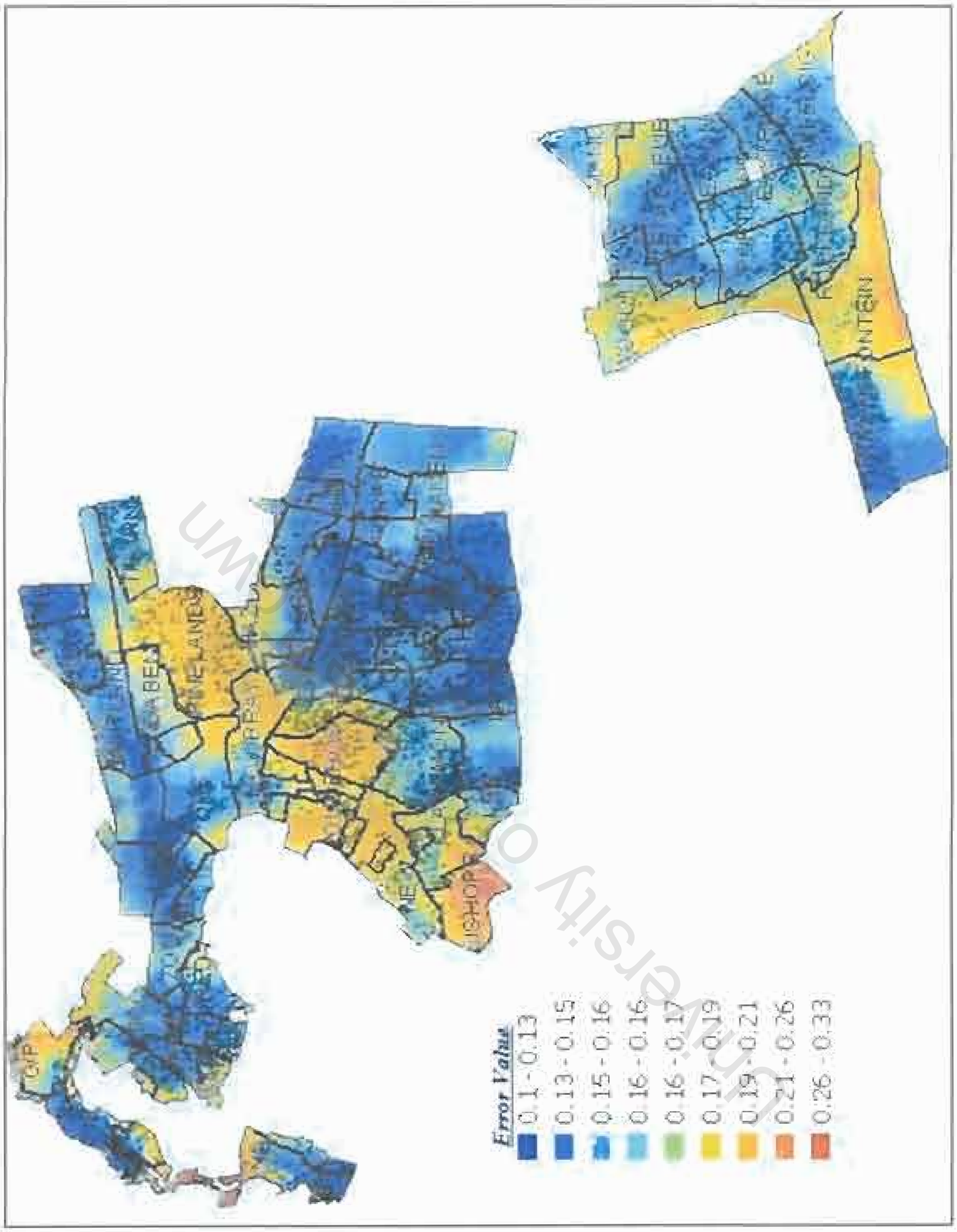


FIGURE 5.9 Error Surface

CHAPTER SIX

Conclusions and Recommendations

This chapter discusses the outcomes of the research and summarises these in the conclusions. Recommendations to the Cape Town City Council (CTCC) are also proposed for future valuations.

6.1 Conclusions

Locational adjustments have been shown in this research to be extremely important if one seeks to produce market value estimates that are representative of market forces, thereby reinforcing the literature, which contends that location is a major variable in computer assisted mass appraisal (CAMA) models. The above statement is supported by the results of this research. All models without location effects fail to produce results that meet the set International Association of Appraisal Officers (IAAO) standards because they fail to capture micro-variations within the neighbourhoods; but once they are incorporated with location factors the appraisal ratios generally improve.

The most suitable technique for generating location factors in the Cape Town MLC is punctual or ordinary kriging, because in all three models: Cape Town (CT), around Table Mountain (ATM) and not around Table Mountain (NATM), it produces results that are consistent and are close to the set IAAO standards. The ability of the ordinary kriging technique to develop a surface based on the statistical relationship between the unknown and known points, results in a surface that is stable and closely resembles the real situation, thereby resulting in location factors that are realistic.

The results produced by the local polynomial interpolator closely follow those of the ordinary kriging technique. The reason for the local polynomial interpolator being better than the global interpolator and inverse distance weighting technique is because the interpolation is based on a number of local polynomials; this enables microvariations with respect to the sale price for each small area to be captured much more easily, compared to the global polynomial technique where a general trend can only be reflected. This is because with the global interpolator, only one polynomial is used for estimation purposes.

The inverse distance weighting technique produces a surface that is very unstable as seen by the high variation of values between appraisal ratios. The high levels of instability probably arise from it being an exact interpolator, which attempts to generate estimates that are replicas of the original data, and so fitting a polynomial to such a dataset is very difficult if not impossible, thereby resulting in a highly unpredictable polynomial.

The value influence centre technique fails to produce satisfactory results, because a single centre of influence is used, whereas in reality there are numerous such centres. Only one centre is used in this research, because if all the centres of influence in the Cape Town MLC had been used, other location modelling techniques would not have been investigated because of time constraints imposed on the research. The trend surface analysis technique fails completely to produce meaningful results because of its high levels of instability. Instability in the sense that computational difficulties arise when the polynomial is raised to the third order and fails to generate a polynomial beyond the third order.

The research also showed that attempting to use geographic stratification for both base model development and locational analysis, did not produce satisfactory results because within these areas (i.e. ATM and NATM), there is a high level of variation between property prices and characteristics. For example properties in Claremont cannot be compared to properties in Silvertown although both neighbourhoods have been placed in the same geographic area (ATM). The best approach for developing base models has been shown to be that of homogenous neighbourhoods as they produce appraisal ratios that closely resemble those set by IAAO.

The potential of the researcher to build high quality models that meet the set IAAO standards is very high, as seen by those developed using homogenous neighbourhoods with location effects. The advantage of building new models is that, it provided the researcher with a better understanding of the concepts behind CAMA, starting from the basic concepts of data analysis through to the recalibration of base models with location factors. This "hands on" approach enabled the researcher to gain confidence in model building, which then lead to further exploration of other techniques for location modelling such as trend surface analysis and homogenous neighbourhoods. The potential to explore different approaches would not have been attempted if base models had been adopted from the CTCC, because the research would have only focused on incorporating location factors into them.

This research reconfirmed the notion that Cape Town's property market is very complex due to the fluctuation of property prices within and between neighbourhoods. The general trend that exists is that property prices pick up as one approaches Table Mountain and the variation becomes worse with properties on the Atlantic Seaboard. The reason for the sharp fluctuations on the Atlantic Seaboard can be attributed to proximity to the beachfront, with properties (e.g. the Clifton bungalows) along the beach selling for exorbitant prices (±R10 000 000) compared to properties situated on the mountainside (±R5 000 000). The variation on the other side of the mountain that is facing the east is better because the only major natural feature is the mountain.

6.2 Recommendations

Scalar values for some qualitative variables generated, based on the dataset for the entire jurisdiction, should be revisited because they tend to have opposite effects for both upmarket and low value neighbourhoods. For example looking at street *(i.e. an expressway, railway station or taxis rank)* it has a major impact in low value neighbourhoods because it provides the transport corridor, whereas in the upmarket neighbourhoods people would rather have their property far away from such places since they are seen as a source of noise pollution and access for criminal activities. Neighbourhoods where such variables are very influential on property prices should be identified and allocated different scalar values to other areas in the jurisdiction.

Neighbourhoods around Table Mountain have been identified as the most problematic when attempting to appraise properties using CAMA techniques. This is due to the high local variation in property prices. The optimal approach that can be used to minimise this problem would be to create micro neighbourhoods *(i.e. subsets of a neighbourhood)* for very high and low value areas within the neighbourhoods. These low value micro neighbourhoods can then be incorporated into other homogenous models, leaving the very high value micro neighbourhoods to form their own homogenous model.

The solution to not having data in some sectors of a neighbourhood can be overcome by site visits by professional valuers so that they can estimate property values also known as proxy sales, which can then be used for development of the response surfaces and overall modelling. This approach will go a long way to minimising the errors at the edges of the jurisdiction, and where no sales have taken place within a specified period.

One major aspect the CTCC Valuation Department have to direct their effort is in data management, as seen in this research where about 32% of the data (i.e. 1901 records from a sample of 6021) had to be rejected either due to miscoding or gross errors. Probably the data manual used by the data collectors should be simplified (i.e. well illustrated) to such an extent that even a person without any property valuation knowledge will be able to apply an objective judgement especially when dealing with very subjective variables such as quality, condition and view.

Blettner, R.A. (1969), "Mass Appraisals via Multiple Regression Analysis", *The Appraisal Journal*, Vol 37, pp 513-521.

Borst, R.A. (1995), "Artificial Neural Networks in Mass Appraisal", *Journal of Property Tax Assessment and Administration*, Vol.1, No.2.

Chilè, J.P. and Delfiner, P. (1999), "Geostatistics: Modeling Spatial Uncertainty", John Wiley and Sons, New York.

Clark, I. (1979), "Practical Geostatistics", Applied Science Publishers Ltd, London.

Davis, J.C. (1973), "Statistics and Data Analysis in Geology", John Wiley and Sons, Canada.

Dorsel, D. and La Breche, T. (1997), "Kriging" , Environmental Sampling and Montiroing, Virginia.
http://www.cee.vt.edu/program_areas/environmental/teach/smprimer/kriging/kriging.html
Last accessed on August 2003.

Dubin, R.A. (1992), "Spatial Autocorrelation and Neighbourhood Quality", *Regional Science and Urban Economics*, Vol.22, pp.433-452.

Dubin, R., Pace, R.K. and Thibodeau, T.G. (1999), "Spatial Autoregression Techniques for Real Estate Data", *Journal of Real Estate Literature*, Vol. 7, pp 79 – 95.

Dubin, R.A. and Sung, C. (1987), "Spatial Variation in the Price of Housing: Rent Gradients in Nonmonocentric Cities", *Urban Studies*, Vol.24, pp.193-204.

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

Eisenlauer, J.F. (1968), "Mass Versus Individual Appraisals", *The Appraisal Journal*, Vol. 36, pp. 532-540.

Figueroa, A.R. (1999), "Modeling the Value of Location in Regina using GIS and Spatial Autocorrelation Statistics", *Assessment Journal*, Vol.6, No.6, pp. 29 – 37.

French, N., Pagourtzi, E., Assimakopoulos, V. and Hatzichristos, T. (2003), "Real Estate Appraisal: A Review of Valuation Methods", *Journal of Property Investment and Finance*, Vol. 21, No. 4, pp. 383 – 401.

Gallimore, P., Fletcher, M. and Carter, M. (1996), "Modeling the Influence of Location on Value", *Journal of Property Valuation and Investment*, Vol.14, No.1, pp. 6 – 19.

Garson, G.D. (2003), "Statsnotes: Regression Analysis", College of Humanities and Social Sciences, North Carolina State University, USA.

Gloudemans, R.J (1999), "Mass Appraisal of Real Property", International Association Of Assessing Officers, Chicago.

Gonzales, M.A.S., Soibelman, L., and Formoso C.T.(2002), "A New Approach to Spatial Analysis in CAMA", *9th European Real Estate Society Conference (ERES2002)*, Glasgow. http://www.inf.unisinos.br/~gonzalez/public.html
Last accessed on June 2003.

Hembd, J. and Infanger, C. (1981), "An Application of Trend Surface Analysis to a Rural-Urban Land Market", *Land Economics*, Vol.57, pp. 303-322.

International Association of Assessing Officers (IAAO), (1999), "Standard on Ratio Studies", *Assessment Journal*, Vol. 6, No. 5, pp. 23 – 65.

Isakson, H. R. (2001), "Using Multiple Regression Analysis in Real Estate Appraisal", *Appraisal Journal*, Vol. 69, No. 4, pp. 424 – 431.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K. and Lucas, N. (2001), "Using ArcGIS Geostatstical Analyst", ESRI, USA.

Kochetkov, Y. (1998), "Error Evaluation in Computer Assisted Mass Appraisal Models, *International Real Estate Conference(ERES&AREUEA 1998)*, Maastricht, The Netherlands. http://www.crea.ru/newcrea/Papers/Eng/err_maas.html
Last accessed on June 2003.

La Rose, T.A. (1988), "Global response surface analysis used to update appraisals in a Computer Assisted Mass Appraisal Environment", *World Congress III of Computer Assisted Valuation and Land Information Systems*, Cambridge, Massachusetts.

Lenk, M., Worzala, E. M. and Silva, A. (1997), "High-Tech Valuation: Should Artificial Neural Networks Bypass the Human Valuer", *Journal of Property Valuation and Investment*, Vol.15, No.1.

Maryland State Department of Assessments and Taxation (SDAT), (2001), 2000 Assessment Ratio Report, USA.
http://www.dat.state.md.us/sdatweb/stats/00rr_rpt.html
Last accessed on June 2003.

Matheron, G. (1963), "Principles of Geostatistics", *Economic Geol*, Vol.58, pp.1246-1266.

McCluskey, W.J. (1996), "Predictive accuracy of machine learning models for the mass appraisal of residential property", *New Zealand Valuers' Journal*, Vol. July, pp. 41-47.

McCluskey, W.J. and Borst, R.A. (1997), "An Evaluation of MRA, Comparable Sales Analysis and ANNs for the Mass Appraisal of Residential Properties in Northern Ireland", *Assessment Journal*, Vol. 4, No. 1, pp. 47 – 58.

McCluskey, W.J., Deddis, W.G., and Lamont, I. (2002), "The Application of Spatially Derived Location Factors Within a GIS Environment", *Eighth Annual PRRES Conference*, Christchurch, New Zealand.
http://business2.unisa.edu.au/prres/Proceedings/2002proceedings.htm
Last accessed on October 2003.

McCluskey, W.J., Deddis, W., McBurney, R.D., Mannis, A. and Borst, R. (1997), "Interactive application of Computer Assisted Mass Appraisal and Geographic Information Systems", *Journal of Property Investment And Finance*, Vol. 15, No. 5, pp. 448 – 465.

O'Connor, M.P. (2002), "Comparison of Three Residential Regression Models: Additive, Multiplicative and Hybrid", *Assessment Journal*, Vol. 9, No. 4, pp. 37 – 43.
Pace, R.K., Barry, R. and Sirmans, C.F. (1998), "Spatial Statistics and Real Estate", *Journal of Real Estate Finance and Economics*, Vol. 17, No. 1, pp. 5 – 13.

Parker, C. (1981), "Trend Surface and the Spatio-Temporal Analysis of Residential Land-use Intensity and Household Housing Expenditures", *Land Economics*, Vol.57, pp. 323-337.

Rummel, R.J. (1976), "Understanding Correlation", Honolulu, Department of Political Science, University of Hawaii, Hawaii.

Scott, I. (1988), "A Knowledge Based Approach to Computer Assisted Mortgage Valuation of Residential Property", Unpublished Phd Thesis, University Of Glamorgan.

Siu, F.L. and Yu, S. (2001), "The sensitivity of the RESET tests to disturbance autocorrelation in regression analysis", *Empirical Economics*, Vol. 26, No. 4, , pp. 721-726.

Swan, A.R.H. and Sandilands, M. (1995), "Introduction to Geological Data Analysis", Blackwell Science Ltd, London.

Tay, D.P. and Ho, D.K. (1994), "Intelligent Mass Appraisal", *Journal of Property Tax Assessment and Administration*, Vol.1, No.1.

Ward, R.W. (2001), "Demonstration of CAMA in South Africa", *Assessment Journal*, Vol. 8, No. 3, pp. 33 – 43.

Ward, R.W. (2004), "Oral Discussion on SPSS and NCSS", Cape Town City Council, South Africa.

Ward, R. W., Guilford, J., Jones, B., Pratt, D. and German, J. (2002), "Piecing Together Location: Three Studies by Lucas County Research and Development Staff", *Assessment Journal*, Vol. 9, No. 5, pp.15 – 48.

Ward, R. W., Ireland, M., Epstein, J. and Field, S. (2000), "Data Analysis Lab", Cape Town City Council, South Africa.

Worzala, E., Lenk, M. and Silva, A. (1995), "An exploration of neural networks and its application to real estate valuation", *The Journal of Real Property Research*, Vol. 10, No. 2.

Agterberg, F.P. (1974), "Developments in Geomathematics 1: Geomathematical Background and Geo-Science Applications", Elsevier Scientific Publishing Company, Amsterdam.

Barbara Curtiss, "Real Property Valuation", Fairfield County Office of the Auditor.
http://www.co.fairfield.oh.us/AUDITOR/VALUE/
Last accessed on May 2003.

Castle, G. (1993), "GIS in Property Valuation – What's It Worth?", *Business Geographics*, Vol. November/December.
http://www.castleconsulting.com/93gisinp.html
Last accessed on May 2003.

Certified Appraisers of Iowa and Nebraska, "Factors of a Professional Real Estate Appraisal".
http://www.certified-appraisers.net/appraisals.html
Last accessed on May 2003.

German, J.C., Robinson, D. and Youngman, J. (2000), "Traditional Methods and New Approaches to Land Valuation, *Land Lines*, Vol. 12, No. 4, pp. 60-77.

Hintze, J.L. (1999), "NCSS 2000 User's Guide-1", Kaysville, Utah.
Kerr, A. W., Hall, H. K. and Kozub, S.A. (2002), "DOING STATISTICS WITH SPSS", SAGE, London.

Petersdorff, T. "Interpolation with Polynomials and Splines", Department of Mathematics, University of Maryland, USA.
http://www.wam.umd.edu/~petersd/interp.html
Last accessed on August 2003.

British Columbia Assessment Authority, "Real Estate Appraisal and Property Assessment", Canada.
http://bcassessment.gov.bc.ca/3_val/3_app.html
Last accessed on May 2003.

Statios, "Reservoir modelling with GSLIB"
http://www.statios.com/Resources/05-kriging.pdf
Last accessed on August 2003.

Thériault M., Kestens, Y. and Des Rosiers, F. (2002), "The Impact of Environmental Perception on Housing Location Choices: A Case Study Among Recent Home Owners in Quebec City", *Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society*, University of Lugano, Switzerland, Vol. 2, pp. 478-483.

Yu, S.M., Siu, K. and Han, S.S. (2000), "Using response surface analysis in Computer Assisted Mass Appraisal to examine the influence of location on property", *The Journal of Property Tax Assessment & Administration*, Vol. 5, No. 4, pp. 3-16.

SINGLE RESIDENTIAL FIELD DATA COLLECTION FORM - GV2000			Unique Ref #
FIELD DATA COLLECTION FORM - GV2000			
Property Ref.	Allotment	Erf No.	Situation Address
Map No.	Suburb	Erf Extent	Property Owner
Price	Sale Date	Category	Zone

1. Existing Use	
Single dwelling residential	1
Other:	2
2. Street	
Expressway	1
Busy	2
Medium	3
Light	4
Dead end	5
3. Off Street Parking	
No	1
Yes	2
4. Dwelling Storeys	
Single storey	1
Single storey + Attic	2
Double storey	3
Double storey + Attic	4
Multilevel	5
5. Dwelling Type	
Sub-Economic	1
Conventional	2
Unconventional	3
Cape Dutch	4
Victorian/Tudor	5
Maisonette	6
Group Housing	7
6. Common Walls	
One	1
Two	2
None	3

7. Roof Covering	
Tile	1
Sheeting	2
Mazista	3
Thatch	4
Other	5
8. External Walls	
Brick	1
Plastered	2
Concrete block	3
Stone	4
Timber	5
Sheeting	6
9. Security	
High	1
Medium	2
Low	3
None	4
10. Garden	
Excellent	1
Above Average	2
Average	3
Below Average	4
None	5
11. View	
Excellent	1
Above Average	2
Average	3
Below Average	4
Poor	5

12. Topography	
Flat	1
Undulating / Uneven	2
Steep	3
13. No. of Rooms	
Living Areas	A
Bedrooms	B
Bathrooms 4+ Fxt	C
Bathrooms 3 Fxt	D
Bathrooms 2 Fxt	E
Water Closet	F
14. Overall Quality	
Excellent	1
Above Average	2
Average	3
Below Average	4
Poor	5
15. General Condition	
Excellent	1
Above Average	2
Average	3
Below Average	4
Poor	5
16. Dwelling Extent	
17. Sales Validity	
Usable improved	1
Usable vacant land	2
Unusable	3
Needs research	4
Changed since sale	5

Yard Items	Extent	AD
18. Covered Stoep		
19. Terrace / Balcony		
20. Deck (wood)		
21. Garage		
22. Car Port		
23. Granny Flat		
24. Servants Qrts		
25. Swimming Pool		
26. Tennis Court*		
27. Shacks*		
28		
29		
30		

A = Attached **D = Detached** *** Provide count only**

31. Visit Record

Date	Time	Insp. Code

32. Inspection Record

Date	Time	Insp. Code	Access Code

Access Codes	
Interior inspection	1
Exterior inspection only	2
No Access to property	3
Information at door	4

33. REMARKS:

OBJECT_ID	

		Model	Test	SALE					SPSM					Scalar
NBHD	**No**	**Sample**	**No**	**Mean**	**Median**	**S.D**	**Min**	**Max**	**Mean**	**Med**	**SD**	**Min**	**Max**	**Value**
Athlone	70	56	14	174909	169711	57090	71190	331470	1586	1523	408	864	2886	0.87
Bantry Bay	13	10	3	1694624	1455700	1040435	820560	4273329	6586	5677	2515	3777	11903	2.00
Beacon Valley	112	90	22	76602	74145	17570	43000	160020	1188	1156	229	674	1675	0.70
Belgravia	38	30	8	186591	172383	73582	90900	420420	1451	1402	373	814	2160	0.85
Belthorn Estate	32	26	6	198187	184944	49329	119485	301840	1655	1766	338	1005	2144	0.93
Bishopscourt	10	8	2	3607503	2857175	1565560	2467500	6858000	8816	8480	2839	5326	13771	2.64
Bridgetown	22	18	4	90779	93840	20393	57860	142020	1605	1665	307	1113	2044	0.83
Camps Bay Bakoven	9	7	2	2035950	2038200	404130	1545600	2681550	13700	14734	4969	6109	20394	2.80
Camps Bay Upper	72	58	14	1419330	1338888	405730	826580	2624950	6504	6380	1463	4140	10004	2.02
Cape Town City / C	4	3	1	449150	499950	106010	327300	520200	4721	5319	1073	3482	5363	1.58
Claremont High	75	60	15	777633	778943	228764	358360	1497250	4195	3981	845	3047	6262	1.57
Claremont Low	235	188	47	362870	349052	102912	182340	867000	2983	3020	459	2001	4031	1.23
Clifton	6	5	1	2854196	2142800	1646995	1671150	5610000	19334	20778	5505	12660	25350	3.17
Crawford	48	38	10	243402	245540	36520	176640	325755	2287	2317	458	1503	3146	1.06
Eastridge	68	54	14	78451	74423	19418	46170	140120	1113	1115	231	624	1743	0.69
Factretown	24	19	5	74207	68380	23551	45239	133380	1235	1235	257	800	1666	0.71
Fresnaye	55	44	11	1225345	1147930	637962	404250	3974400	5100	4768	1953	1831	9092	1.80
Gardens Lower	46	37	9	507381	352000	156419	202335	798680	3234	3103	677	2140	4930	1.35
Gardens Upper	27	22	5	1172163	1114700	329589	658450	1940400	5244	4791	1854	2575	9751	1.79
Gatesville	11	9	2	211425	192200	55572	122475	291165	1972	2045	476	1387	2795	0.98
Green Point	48	38	10	638657	599370	246265	266760	1249300	4039	3823	1373	1818	8523	1.46
Hanover Park	23	18	5	63670	63903	13828	33000	91170	985	1018	204	580	1357	0.66
Hazendal	7	6	1	176302	189886	81459	80000	279250	1507	1477	135	1321	1672	0.88
Heideveld	11	9	2	66898	66240	13752	42772	83904	1224	1244	240	807	1583	0.71
Kenilworth High	23	18	5	917207	933093	290701	375700	1487700	3569	3538	869	2339	5189	1.59
Kenilworth Low	78	62	16	338452	334560	104228	131264	574560	2883	2828	478	1948	4228	1.19
Kensington	34	27	7	173707	174879	35126	101420	245000	1564	1569	344	1004	2534	0.88
Kewtown	19	15	4	82707	84132	20728	48510	140140	1573	1700	302	1039	2020	0.83
Lansdowne High	160	128	32	223273	222160	45679	128250	367250	2088	2070	381	1390	3296	1.01
Lansdowne Low	50	40	10	202198	193263	48002	120015	309060	1807	1800	300	1245	2443	0.94
Lentegeur	130	104	26	77722	77115	13264	44404	135600	1194	1180	234	622	1756	0.71
Maitland	22	18	4	157091	157540	27042	103645	220350	1374	1295	267	1065	1950	0.81
Mandalay	50	40	10	134445	132480	28949	81040	204820	1632	1517	521	1031	4284	0.84
Mouille Point	1	1	0	401488	401488		401488	401488	2996	2996		2996	2996	1.26
Mountview	7	6	1	200702	166910	106471	92200	361600	1490	1587	351	1002	1817	0.88
Mowbray	60	48	12	305529	307486	104471	115920	511200	2324	2345	465	1415	3680	1.11
Ndabeni	6	5	1	190831	192465	57896	117285	270180	1562	1580	184	1283	1758	0.90
Newfields	5	4	1	152092	148845	56106	88395	222283	1413	1393	441	1004	1860	0.82
Newlands	92	74	18	865747	797875	318750	300000	1932000	4626	4428	904	3134	7051	1.62
Observatory	150	120	30	266579	251192	86311	93598	474725	2352	2332	470	1232	3631	1.07
Oranjezicht Lower	24	19	5	569900	544491	146734	290000	1011750	3317	3370	604	2368	4530	1.38
Oranjezicht Upper	28	22	6	1073940	1052035	262317	529800	1718325	4336	4328	830	2731	6026	1.72
Parktown	2	2	0	67308	67308	26584	48510	86105	985	985	230	822	1148	0.65
Penlyn Estate	10	8	2	218639	188325	66289	155850	319500	1560	1485	268	1231	2065	0.88
Pinati	15	12	3	177514	173580	66121	63900	302895	1459	1465	220	1022	1770	0.86
Pinelands	232	186	46	452497	420798	136392	185322	988750	2899	2919	480	1986	4696	1.26
Portland	146	117	29	100611	98260	18584	49095	149040	1300	1288	228	768	1783	0.76
Primrose Park	4	3	1	141677	148000	40695	98190	178840	1671	1806	364	1259	1947	0.90
Rocklands	145	116	29	81188	80885	13675	35171	124848	1266	1275	211	639	1713	0.74
Rondebosch East	60	48	12	277529	266805	57921	187680	518880	2215	2180	313	1587	2951	1.05
Rondebosch High	107	86	21	801923	730900	345525	319095	1924700	3816	3877	666	2105	5344	1.53
Rondebosch Low	120	96	24	499390	493853	140842	191007	850000	3035	3015	550	1784	4331	1.31
Rosebank Middle	12	10	2	267498	271578	53434	177660	373460	2570	2431	663	1734	4047	1.10
Rosebank Upper 1	3	2	1	645583	645583	304816	430045	861120	3115	3115	351	2867	3364	1.40
Rosebank Upper 2	12	10	2	595927	534980	173111	359939	935100	3003	2847	470	2522	3805	1.31
Rylands	19	15	4	270933	241960	144569	125730	663600	1639	1555	449	1083	2638	0.93
Salt River High	11	9	2	236410	235980	48209	157015	291060	2479	2457	351	1807	2982	1.08
Salt River Low	13	10	3	148544	142128	29914	114790	205740	1593	1593	321	1159	2143	0.86
Schotschekloof	13	10	3	237855	230673	68191	156380	368200	2345	2351	595	1528	3750	1.06
Sea Point	74	59	15	681569	560000	393588	180000	2275000	3749	3481	1187	1989	7225	1.40
Silvertown	22	18	4	97829	91825	26965	49248	150920	1504	1515	316	978	2020	0.80

		Model	Test	SALE					SPSM					Scalar
NBHD	No	Sample	No	Mean	Median	S.D	Min	Max	Mean	Med	SD	Min	Max	Value
Strandfontein East	196	157	39	136375	133125	25720	65412	217170	1608	1614	237	967	2215	0.86
Strandfontein West	57	46	11	116243	117150	22523	74538	165735	1899	1863	320	1147	2467	0.89
Surrey	21	17	4	192757	194940	34585	98700	247940	1738	1732	515	864	2618	0.93
Sybrand Park	30	24	6	268573	262826	63537	145695	473960	2014	1934	350	1275	2609	1.01
Tafelsig	89	71	18	65321	64032	10165	42080	94945	978	985	175	607	1417	0.65
Tamboerskloof	47	38	9	760815	715100	233808	398160	1299500	4051	3889	754	2706	6159	1.53
Thornton	93	74	19	234733	230798	36363	121560	362100	1984	1997	288	1465	2656	1.00
Three Anchor Bay	13	10	3	506026	457250	150320	328320	843880	3049	3036	768	2091	4489	1.29
Vanguard	15	12	3	179530	165975	39613	110400	243120	1539	1469	293	1121	2078	0.85
Vredehoek	62	50	12	535572	504920	158882	287280	1196800	3575	3706	717	2085	4953	1.40
Welcome	6	5	1	176864	180000	52679	101175	246240	1389	1483	291	992	1652	0.87
Weltevreden Valley	107	86	21	113470	109888	23427	76950	181825	1871	1833	297	1348	2770	0.88
Westridge	137	110	27	102950	101714	17326	71820	149842	1291	1274	275	795	2206	0.76
Woodlands	80	64	16	91633	86028	23542	53459	175150	1226	1224	222	731	1952	0.73
Woodstock High	27	22	5	329405	324235	104667	146970	532500	2713	2902	660	1517	3825	1.20
Woodstock Low	44	35	9	127299	121440	37114	65250	204240	1471	1447	344	837	2258	0.81
Woodstock Middle	60	48	12	205653	191523	59679	105192	361600	2221	2171	502	1215	3503	1.00
Zonnebloem	11	9	2	277948	286230	51345	203148	371770	3026	2663	707	2156	4323	1.14

			SALE		SPSM		Scalar	
		No	Mean	Median	Mean	Median	Value	Frequency Graph
STR								
	1	43	266339	231977	2224	2137	0.80	
	2	296	335634	141581	2281	1674	0.86	
	3	473	301011	183260	2252	1900	0.94	
	4	2212	345554	226878	2415	2093	1.00	
	5	275	533643	293770	3393	2649	1.13	
PARK								
	1	293	321920	226380	2811	2433	0.93	
	2	3006	355946	218325	2425	2000	1.00	
SEC								
	1	176	956731	769163	4585	3955	1.75	
	2	631	607826	462400	3400	3140	1.46	
	3	2134	246165	163646	2080	1758	1.00	
	4	358	243178	134348	2016	1641	0.95	
TOPO								
	1	3030	306261	203901	2301	1949	1.00	
	2	117	694228	505875	3334	3010	1.36	
	3	152	1020393	951185	4925	4491	1.76	
VIEW								
	1	59	1666172	1545600	7567	6079	2.22	
	2	343	814789	709100	4188	3901	1.63	
	3	2768	269193	187035	2136	1862	1.00	
	4	116	320597	258820	2440	2240	0.94	
	5	13	323290	296100	2549	2265	0.88	
GARD								
	1	9	1664108	1455825	5761	4370	1.45	
	2	130	1198446	992533	4784	4342	1.32	
	3	850	555517	429400	3249	3003	1.00	
	4	629	332068	248678	2502	2161	0.82	
	5	1681	185878	114790	1846	1530	0.65	
DTYPE								
	1	297	78062	73872	1260	1231	0.69	
	2	2707	371544	244188	2585	2207	1.00	
	3	36	1497153	1307775	5180	4749	1.79	
	4	11	882112	750000	3892	3618	1.45	
	5	70	549761	514000	3184	3003	1.27	
	6	166	176733	86566	1524	1083	1.20	
	7	12	326653	318320	2850	3291	1.20	
CWALL								
	1	906	188657	107996	1936	1517	0.83	
	2	263	199942	145210	2084	1756	0.75	
	3	2130	441685	285725	2728	2285	1.00	
COND								
	1	21	1091784	999940	5100	4739	1.79	
	2	208	1038306	874100	4711	4128	1.68	
	3	2792	306905	206386	2314	1968	1.00	
	4	265	250314	168165	2062	1826	0.94	
	5	13	168372	87126	1451	1374	0.78	
QUAL								
	1	14	1382700	1474725	5895	5325	1.95	
	2	243	1070407	916737	4648	4270	1.63	
	3	2615	320544	231770	2418	2132	1.00	
	4	427	109150	77280	1351	1232	0.70	
STOREYS								
	1	2790	299005	202478	2360	1973	1.00	
	2	56	685683	584370	3332	3181	1.32	
	3	419	611182	506480	2859	2705	1.43	
	4	6	2075855	1172425	5868	4261	1.84	
	5	8	1335004	1341000	5619	5377	2.01	

<table>
<tr><th></th><th></th><th colspan="2">SALE</th><th colspan="2">SPSM</th><th>Scalar</th><th></th></tr>
<tr><th></th><th>No</th><th>Mean</th><th>Median</th><th>Mean</th><th>Median</th><th>Value</th><th>Frequency Graph</th></tr>
<tr><td>ROOFING</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1</td><td>2185</td><td>371740</td><td>240155</td><td>2496</td><td>2113</td><td>1.00</td><td>||||||||||||||||||||||</td></tr>
<tr><td>2</td><td>976</td><td>241033</td><td>161830</td><td>2158</td><td>1711</td><td>0.87</td><td>||||||||||</td></tr>
<tr><td>3</td><td>56</td><td>974660</td><td>761578</td><td>4084</td><td>3496</td><td>1.47</td><td>|</td></tr>
<tr><td>4</td><td>13</td><td>623377</td><td>592230</td><td>3280</td><td>3451</td><td>1.38</td><td>|</td></tr>
<tr><td>5</td><td>49</td><td>850135</td><td>800030</td><td>4467</td><td>4463</td><td>1.59</td><td>|</td></tr>
<tr><td>EXWALLS</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1</td><td>382</td><td>233604</td><td>95398</td><td>1868</td><td>1441</td><td>0.73</td><td>||||</td></tr>
<tr><td>2</td><td>2580</td><td>391654</td><td>262945</td><td>2630</td><td>2305</td><td>1.00</td><td>||||||||||||||||||||||||||</td></tr>
<tr><td>3</td><td>299</td><td>89295</td><td>75999</td><td>1232</td><td>1182</td><td>0.66</td><td>|||</td></tr>
<tr><td>4</td><td>5</td><td>507130</td><td>757100</td><td>2638</td><td>3277</td><td>1.38</td><td>|</td></tr>
<tr><td>5</td><td>13</td><td>1761712</td><td>1723780</td><td>12987</td><td>12660</td><td>2.42</td><td>|</td></tr>
</table>

Linearised Variable	Transformation
STR_ADJ	LOOKUP (STR, STRCODE, STRLIN, 1)
PARK_ADJ	LOOKUP (PARK, PARKCODE, PARKLIN, 1)
SEC_ADJ	LOOKUP (SEC, SECCODE, SECLIN, 1)
TOPO_ADJ	LOOKUP (TOPO, TOPOCODE, TOPOLIN, 1)
VIEW_ADJ	LOOKUP (VIEW, VIEWCODE, VIEWLIN, 1)
GARD_ADJ	LOOKUP (GARD, GARDCODE, GARDLIN, 1)
DTYPE_ADJ	LOOKUP (DTYPE, DTYPECODE, DTYPELIN, 1)
CWALL_ADJ	LOOKUP (CWALL, CWALLCODE, CWALLLIN, 1)
COND_ADJ	LOOKUP (COND, CONDCODE, CONDLIN, 1)
QUAL_ADJ	LOOKUP (QUAL, QUALCODE, QUALLIN, 1)
STOREYS_ADJ	LOOKUP (STOREYS, STOREYSCODE, STOREYSLIN, 1)
ROOFING_ADJ	LOOKUP (ROOFING, ROOFINGCODE, ROOFINGLIN, 1)
EXWALLS_ADJ	LOOKUP (EXWALLS, EXWALLSCODE, EXWALLSLIN, 1)
ATM	LOOKUP (NBHD, ATMcode, ATMlin, 1)
NBHD_ADJ	LOOKUP (NBHD, NBHDCODE, NBHDLIN, 1)
GROUP	LOOKUP (NBHD, GROUP_Code, GROUP_Lin, 1)
Location Factor	LOOKUP (PID, PIDx, PREDICTEDX, 0)

	SALE	STR	PARK	SEC	TOPO	VIEW	GARD	DTYPE	CWALL	CONI	QUAI	POOL	TLA	TOTFIX	GFSQ	GARCP	STOREYS	ROOF	EXWAI	ERFEXT
SALE	1	0.11	0.02	0.48	0.39	0.57	0.56	0.30	0.28	0.46	0.54	0.40	0.79	0.64	0.15	0.44	0.32	0.29	0.31	0.63
STR	0.11	1	-0.02	0.02	0.07	0.09	0.04	0.03	0.03	0.09	0.08	0.01	0.06	0.06	0.00	0.02	0.00	0.03	0.11	0.02
PARK	0.02	-0.02	1	0.02	-0.04	-0.04	0.10	-0.09	0.36	0.06	0.06	0.11	0.06	0.05	0.06	0.27	-0.02	0.11	-0.11	0.21
SEC	0.48	0.02	0.02	1	0.18	0.25	0.47	0.16	0.18	0.34	0.39	0.28	0.49	0.45	0.08	0.32	0.19	0.20	0.16	0.30
TOPO	0.39	0.07	-0.04	0.18	1	0.48	0.18	0.15	0.04	0.16	0.21	0.11	0.28	0.20	0.09	0.16	0.16	0.11	0.06	0.18
VIEW	0.57	0.09	-0.04	0.25	0.48	1	0.28	0.17	0.11	0.29	0.32	0.17	0.40	0.29	0.09	0.23	0.23	0.12	0.22	0.23
GARD	0.56	0.04	0.10	0.47	0.18	0.28	1	0.22	0.37	0.33	0.45	0.39	0.62	0.52	0.17	0.48	0.15	0.27	0.22	0.56
DTYPE	0.30	0.03	-0.09	0.16	0.15	0.17	0.22	1	0.11	0.18	0.35	0.10	0.35	0.26	0.05	0.17	0.38	0.14	0.19	0.18
CWALL	0.28	0.03	0.36	0.18	0.04	0.11	0.37	0.11	1	0.14	0.29	0.27	0.38	0.28	0.14	0.43	-0.11	0.23	0.29	0.47
COND	0.46	0.09	0.06	0.34	0.16	0.29	0.33	0.18	0.14	1	0.58	0.20	0.37	0.36	0.09	0.25	0.21	0.19	0.09	0.23
QUAL	0.54	0.08	0.06	0.39	0.21	0.32	0.45	0.35	0.29	0.58	1	0.31	0.52	0.46	0.12	0.38	0.19	0.29	0.27	0.38
POOL	0.40	0.01	0.11	0.28	0.11	0.17	0.39	0.10	0.27	0.20	0.31	1	0.50	0.41	0.12	0.38	0.12	0.19	0.09	0.49
TLA	0.79	0.06	0.06	0.49	0.28	0.40	0.62	0.35	0.38	0.37	0.52	0.50	1	0.72	0.18	0.53	0.39	0.32	0.25	0.67
TOTFIX	0.64	0.06	0.05	0.45	0.20	0.29	0.52	0.26	0.28	0.36	0.46	0.41	0.72	1	0.14	0.43	0.23	0.24	0.24	0.52
GFSQ	0.15	0.00	0.06	0.08	0.09	0.09	0.17	0.05	0.14	0.09	0.12	0.12	0.18	0.14	1	0.16	0.02	0.10	0.07	0.23
GARCP	0.44	0.02	0.27	0.32	0.16	0.23	0.48	0.17	0.43	0.25	0.38	0.38	0.53	0.43	0.16	1	0.11	0.25	0.19	0.54
STOREYS	0.32	0.00	-0.02	0.19	0.16	0.23	0.15	0.38	-0.11	0.21	0.19	0.12	0.39	0.23	0.02	0.11	1	0.12	-0.09	0.11
ROOFING	0.29	0.03	0.11	0.20	0.11	0.12	0.27	0.14	0.23	0.19	0.29	0.19	0.32	0.24	0.10	0.25	0.12	1	0.11	0.29
EXWALLS	0.31	0.11	-0.11	0.16	0.06	0.22	0.22	0.19	0.29	0.09	0.27	0.09	0.25	0.24	0.07	0.19	-0.09	0.11	1	0.19
ERFEXT	0.63	0.02	0.21	0.30	0.18	0.23	0.56	0.18	0.47	0.23	0.38	0.49	0.67	0.52	0.23	0.54	0.11	0.29	0.19	1

			SALE APPRAISAL RATIO		
Sale Month	Month	Year	No of Sales	Mean	Std Dev
1	January	1999	174	0.848	0.336
2	February	1999	225	0.809	0.325
3	March	1999	246	0.830	0.298
4	April	1999	196	0.814	0.304
5	May	1999	226	0.845	0.308
6	June	1999	165	0.892	0.346
7	July	1999	223	0.845	0.309
8	August	1999	251	0.897	0.353
9	September	1999	218	0.867	0.285
10	October	1999	261	0.922	0.311
11	November	1999	238	0.904	0.326
12	December	1999	163	0.944	0.335
13	January	2000	186	0.998	0.362
14	February	2000	203	1.008	0.374
15	March	2000	95	1.060	0.398
16	April	2000	25	1.089	0.357
17	May	2000	27	1.066	0.653
18	June	2000	44	1.160	0.584
19	July	2000	42	1.006	0.452
20	August	2000	33	1.046	0.409
21	September	2000	25	0.859	0.335
22	October	2000	17	0.957	0.560
23	November	2000	14	0.822	0.438
24	December	2000	2	1.234	0.410

Dependent *Sale Price*

Iteration Detail Section

Iteration 0: *Unchanged*

In	Variable	Standard. Coefficient	R-Squared Increment	R-Squared Other X's	T-Value	Prob Level	Pct Change Sqrt (MSE)
Yes	Str	0.030	0.0009	0.026	3.52	0.000	0.175
Yes	Park	-0.008	0.0000	0.235	-0.79	0.432	-0.006
Yes	Sec	0.067	0.0030	0.335	6.50	0.000	0.630
Yes	Topo	0.072	0.0038	0.257	7.33	0.000	0.806
Yes	View	0.230	0.0340	0.359	21.93	0.000	7.111
Yes	Gard	0.024	0.0003	0.492	2.04	0.041	0.049
Yes	Dtype	0.001	0.0000	0.278	0.09	0.932	-0.015
Yes	Cwall	-0.065	0.0025	0.414	-5.92	0.000	0.520
Yes	Cond	-0.089	0.0004	0.949	-2.37	0.018	0.071
Yes	Cond_Qual	0.369	0.0022	0.983	5.64	0.000	0.472
Yes	Qual	-0.179	0.0015	0.954	-4.54	0.000	0.301
Yes	Pool	-0.019	0.0002	0.311	-1.84	0.066	0.037
Yes	Tla	0.358	0.0316	0.754	21.14	0.000	6.623
Yes	Totfix	0.088	0.0035	0.549	7.01	0.000	0.736
Yes	Gfsq	-0.023	0.0005	0.061	-2.64	0.008	0.092
Yes	Garcp	-0.047	0.0013	0.420	-4.28	0.000	0.266
Yes	Storeys	0.034	0.0007	0.345	3.24	0.001	0.145
Yes	Roofing	0.013	0.0001	0.146	1.39	0.165	0.014
Yes	Exwalls	0.094	0.0067	0.240	9.78	0.000	1.441
Yes	Erfext	0.254	0.0271	0.579	19.59	0.000	5.710
R-Squared = 0.769924 Sqrt (MSE) = 199221.5							

Iteration 1: *Removed DTYPE from equation*

In	Variable	Standard. Coefficient	R-Squared Increment	R-Squared Other X's	T-Value	Prob Level	Pct Change Sqrt (MSE)
Yes	Str	0.030	0.0009	0.026	3.52	0.000	0.175
Yes	Park	-0.008	0.0000	0.227	-0.80	0.424	-0.006
Yes	Sec	0.067	0.0030	0.333	6.50	0.000	0.632
Yes	Topo	0.072	0.0038	0.256	7.35	0.000	0.809
Yes	View	0.230	0.0340	0.357	21.95	0.000	7.124
Yes	Gard	0.024	0.0003	0.492	2.05	0.041	0.049
Yes	Cwall	-0.065	0.0025	0.412	-5.92	0.000	0.521
Yes	Cond	-0.088	0.0004	0.949	-2.37	0.018	0.071
Yes	Cond_Qual	0.368	0.0023	0.983	5.66	0.000	0.474
Yes	Qual	-0.178	0.0015	0.953	-4.58	0.000	0.305
Yes	Pool	-0.019	0.0002	0.308	-1.85	0.064	0.037
Yes	Tla	0.359	0.0318	0.753	21.21	0.000	6.661
Yes	Totfix	0.088	0.0035	0.549	7.02	0.000	0.737
Yes	Gfsq	-0.023	0.0005	0.061	-2.64	0.008	0.092
Yes	Garcp	-0.047	0.0013	0.420	-4.28	0.000	0.266
Yes	Storeys	0.034	0.0008	0.275	3.44	0.001	0.166
Yes	Roofing	0.013	0.0001	0.146	1.39	0.165	0.014
Yes	Exwalls	0.094	0.0068	0.231	9.84	0.000	1.460
Yes	Erfext	0.254	0.0271	0.579	19.60	0.000	5.711
No	Dtype		0.0000	0.278	0.09	0.932	0.015
R-Squared = 0.769923 Sqrt (MSE) = 199191.2							

Iteration 2: *Removed PARK from equation*

In	Variable	Standard. Coefficient	R-Squared Increment	R-Squared Other X's	T-Value	Prob Level	Pct Change Sqrt (MSE)
Yes	Str	0.030	0.0009	0.026	3.52	0.000	0.174
Yes	Sec	0.067	0.0030	0.333	6.51	0.000	0.632
Yes	Topo	0.072	0.0038	0.255	7.38	0.000	0.816
Yes	View	0.230	0.0341	0.357	21.98	0.000	7.135
Yes	Gard	0.024	0.0003	0.492	2.07	0.039	0.050
Yes	Cwall	-0.068	0.0030	0.343	-6.53	0.000	0.637
Yes	Cond	-0.088	0.0004	0.949	-2.36	0.018	0.070
Yes	Cond_Qual	0.367	0.0022	0.983	5.64	0.000	0.471
Yes	Qual	-0.177	0.0015	0.953	-4.55	0.000	0.302
Yes	Pool	-0.019	0.0002	0.308	-1.86	0.063	0.038
Yes	Tla	0.360	0.0326	0.749	21.48	0.000	6.829
Yes	Totfix	0.088	0.0035	0.549	7.01	0.000	0.736
Yes	Gfsq	-0.023	0.0005	0.061	-2.65	0.008	0.092
Yes	Garcp	-0.049	0.0014	0.401	-4.51	0.000	0.295
Yes	Storeys	0.034	0.0008	0.273	3.40	0.001	0.162
Yes	Roofing	0.012	0.0001	0.145	1.36	0.175	0.013
Yes	Exwalls	0.096	0.0074	0.197	10.23	0.000	1.577
Yes	Erfext	0.253	0.0271	0.576	19.60	0.000	5.710
No	Park		0.0000	0.227	0.80	0.424	0.006
No	Dtype		0.0000	0.270	0.17	0.867	0.015
R-Squared = 0.769878 Sqrt (MSE) = 199180.2							

Iteration 3: *Unchanged*

In	Variable	Standard. Coefficient	R-Squared Increment	R-Squared Other X's	T-Value	Prob Level	Pct Change Sqrt (MSE)
Yes	Str	0.030	0.0009	0.026	3.52	0.000	0.174
Yes	Sec	0.067	0.0030	0.333	6.51	0.000	0.632
Yes	Topo	0.072	0.0038	0.255	7.38	0.000	0.816
Yes	View	0.230	0.0341	0.357	21.98	0.000	7.135
Yes	Gard	0.024	0.0003	0.492	2.07	0.039	0.050
Yes	Cwall	-0.068	0.0030	0.343	-6.53	0.000	0.637
Yes	Cond	-0.088	0.0004	0.949	-2.36	0.018	0.070
Yes	Cond_Qual	0.367	0.0022	0.983	5.64	0.000	0.471
Yes	Qual	-0.177	0.0015	0.953	-4.55	0.000	0.302
Yes	Pool	-0.019	0.0002	0.308	-1.86	0.063	0.038
Yes	Tla	0.360	0.0326	0.749	21.48	0.000	6.829
Yes	Totfix	0.088	0.0035	0.549	7.01	0.000	0.736
Yes	Gfsq	-0.023	0.0005	0.061	-2.65	0.008	0.092
Yes	Garcp	-0.049	0.0014	0.401	-4.51	0.000	0.295
Yes	Storeys	0.034	0.0008	0.273	3.40	0.001	0.162
Yes	Roofing	0.012	0.0001	0.145	1.36	0.175	0.013
Yes	Exwalls	0.096	0.0074	0.197	10.23	0.000	1.577
Yes	Erfext	0.253	0.0271	0.576	19.60	0.000	5.710
No	Park		0.0000	0.227	0.80	0.424	0.006
No	Dtype		0.0000	0.270	0.17	0.867	0.015
R-Squared = 0.769878 Sqrt (MSE) = 199180.2							

Around Table Mountain	Not Around Table Mountain			Black Local Authority
Bantry Bay	Athlone	Ndabeni	Tafelsig	Crossroads
Bishopscourt	Beacon Valley	Newfields	Thornton	Drift Sands
Camps Bay Bakoven	Belgravia	Observatory	Three Anchor Bay	Dunoon
Camps Bay Upper	Belthorn Estate	Parktown	Vanguard	Guguletu
Cape Town City Centre	Bridgetown	Penlyn Estate	Vredehoek	Guguletu N
Claremont High	Claremont Low	Pinati	Welcome	Guguletu S
Clifton	Crawford	Pinelands	Weltevreden Valley	Imizamo Yethu
Constantia	Eastridge	Portland	Westridge	Khayelitsha
Constantia High	Factretown	Primrose Park	Woodlands	Khayelitsha Hi
Fresnaye	Gatesville	Rocklands	Woodstock High	Kraaifontein East
Gardens Lower	Hanover Park	Rondebosch East	Woodstock Low	Langa
Gardens Upper	Hazendal	Rondebosch Low	Woodstock Middle	Langa Sth
Green Point	Heideveld	Rosebank Middle	Zonnebloem	Mfuleni
Kenilworth High	Kenilworth Low	Rosebank Upper 1		Nomzamo Lwandle
Llandudno	Kensington	Rylands		Nyanga
Mouille Point	Kewtown	Salt River High		Nyanga S
Newlands	Lansdowne High	Salt River Low		Philippi
Oranjezicht Lower	Lansdowne Low	Schotschekloof		Sir Lowry's Pass
Oranjezicht Upper	Lentegeur	Silvertown		
Rondebosch High	Maitland	Strandfontein East		
Rosebank Upper 2	Mandalay	Strandfontein West		
Sea Point	Mountview	Surrey		
Tamboerskloof	Mowbray	Sybrand Park		

Selected Method	**Ordinary Kriging**	
Output	**Prediction Map**	
	Surface based on variables	**Surface based on SP**
No of datasets	1	1
No of Points	3299	3299
Order of Trend Removal	Second (Estimated by Global Polynomial Interpolation)	
Transformation		Log
Semivariogram/Covariance		
Model	0.011397*Gaussian (697.32,442.21,320.2) +0.030504*Nugget	0.0058475*Gaussian (697.32,579.84,350.6) +0.019387*Nugget
Error modeling:		
Microstructure	0.030504 (100%)	0.019387 (100%)
Measurement error	0 (0%)	0 (0%)
Searching Neighborhood:		
Neighbors to Include	4 or at least 2 for each angular sector	
Searching Ellipse:		
Angle	305	305
Major Semi axis	100	100
Minor Semi axis	50	50
Angular Sectors	4	4

City of Cape Town Model

	Modelling Sample		Test Sample	
Variable Name	Parameter	T-Value	Parameter	T-Value
	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)
View	0.62	22.34	0.91	19.66
Cond_Qual	0.26	12.40	0.44	9.01
Dtype	0.21	4.79	0.95	12.16
Cwall	0.76	5.37	-0.02	-0.08
Storeys	0.19	4.02	0.05	0.48
Roofing	0.18	3.57	-0.82	-8.31
Exwalls	0.95	23.67	1.16	12.81
Tla	1747.88	28.52	1226.88	10.05
Totfix	17036.61	13.31	12499.80	8.83
Topo	1.46	8.69	1.44	9.64
Sec	2.71	10.23	1.00	4.13
Erfext	47.99	6.03	144.66	6.12
R-Squared		0.79		0.91
Iterations		11		15

	Sale			Sale		
	Actual	Predicted	Ratio	Actual	Predicted	Ratio
No of Cases	3279	3279	3279	812	812	812
Mean	351278	363657	1.25	360455.2	364042.8	1.19
Median	218592	268039.8	1.17	223122.5	256867.7	1.13
Minimum	33000	46404.89	0.24	35958	45369.64	0.36
Lower Quartile	107350	160610.7	0.92	107765	157381.2	0.88
Upper Quartile	426000	433970.1	1.50	425040	401695.4	1.41
Maximum	6858000	5099925	4.90	8280000	8456541	3.57
Range	6825000	5053520	4.67	8244042	8411172	3.20
IQR	318650	273359.4	0.57	317275	244314.2	0.53
Variance	1.71E+11	1.26E+11	0.21	2.87E+11	2.58E+11	0.18
Std. Deviation	414068	354656.4	0.46	535286.7	507764.3	0.43
CoV	1.18	0.98	0.37	1.49	1.39	0.36
Ave |Dev. from Median|	232757	204325.5	0.35	241587.4	211051.4	0.33
CoD	106.48	76.23	29.96	108.28	82.16	28.77
Weighted Mean			1.04			1.01
PRD			120.6			117.9

Around Table Mountain Model

	Modelling Sample		Test Sample	
Variable	Parameter	T-Value	Parameter	T-Value
Name	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)
Str	0.45	2.21	1.53	4.21
View	0.44	7.97	0.73	6.12
Cond	0.17	4.65	0.59	4.87
Exwalls	0.80	9.89	1.33	6.12
Tla	2370.73	13.21	1393.76	3.66
Totfix	23727.52	5.93	11522.95	2.18
Sec	1.39	3.65	-0.27	-0.64
Topo	0.90	2.63	0.99	4.36
Erfext	114.99	3.98	416.63	4.20
R-Squared		0.64		0.85
Iterations		28		16

	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted	
No of Cases	611	611	611	149	149	149
Mean	945256	935073.6	1.05	997661	960811	0.96
Median	800000	814803.8	1.02	813600	712571.1	0.89
Minimum	180000	197264.7	0.30	248940	168574.2	0.42
Lower Quartile	571795	597133.3	0.85	542850	456019.8	0.73
Upper Quartile	1135050	1119139	1.22	1128600	1179904	1.16
Maximum	6858000	5109022	2.94	8280000	8288340	2.06
Range	6678000	4911758	2.64	8031060	8119766	1.64
IQR	563255	522005.9	0.37	585750	723883.6	0.43
Variance	3.86E+11	2.67E+11	0.09	9.59E+11	8.91E+11	0.10
Std. Deviation	621193.3	516339.4	0.30	9.79E+05	9.44E+05	0.32
CoV	0.657169	0.55	0.29	0.98	0.98	0.33
Ave |Dev. from Median|	380915.2	348681.6	0.23	450217.8	506827.6	0.25
CoD	47.6144	42.79	22.09	55.34	71.13	28.08
Weighted Mean			0.99			0.96
PRD			106.6			99.2

Not Around Table Mountain Model

	Modelling Sample			Test Sample		
Variable	Parameter		T-Value	Parameter		T-Value
Name	Estimate		(Ho: Bi=0)	Estimate		(Ho: Bi=0)
View	0.50		14.41	0.82		11.50
Cond_Qual	0.11		3.14	0.39		6.03
Dtype	0.59		8.57	0.19		1.50
Exwalls	0.71		8.86	0.82		6.45
Tla	1200.37		25.85	1414.83		14.61
Totfix	8643.20		11.79	6713.28		4.48
Str	1.73		4.77	2.49		2.94
Sec	1.50		12.05	0.68		3.26
Erfext	119.51		12.01	136.22		7.31
Gard	1.65		3.74	-17.45		-0.22
Pool	1063.30		6.27	0.02		0.03
Garcp	627.53		4.88	-0.57		-0.03
R-Squared			0.78			0.78
Iterations			22			50
	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted	
No of Cases	2668	2668	2668	663	663	663
Mean	215251.1	219446.2	1.15	217252.2	223139.3	1.17
Median	168310	187319.4	1.10	172725	196793.9	1.12
Minimum	33000	49936.57	0.46	35958	52050.21	0.45
Lower Quartile	94945	125252.9	0.88	96050	126259.7	0.91
Upper Quartile	289300	283553.7	1.36	284040	286665.6	1.40
Maximum	1196800	1074012	3.13	1233750	984475.4	3.08
Range	1163800	1024075	2.67	1197792	932425.2	2.63
IQR	194355	158300.8	0.48	187990	160405.9	0.49
Variance	2.31E+10	1.62E+10	0.13	2.46E+10	1.66E+10	0.14
Std. Deviation	152099	127392	0.36	156878.6	128851.5	0.37
CoV	0.71	0.58	0.31	0.72	0.58	0.32
Ave |Dev. from Median|	113801.3	94451.21	0.28	116032.6	94593.27	0.29
CoD	67.61	50.42	25.17	67.18	48.07	25.87
Weighted Mean			1.02			1.03
PRD			113.0			114.4

City of Cape Town Model

	GLOBAL				IDW			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	0.79	36.85	0.94	33.19	1.17	0.00	0.96	33.86
View	0.25	10.25	0.65	20.99	0.00	0.00	0.32	9.11
Cond_Qual	0.16	9.00	0.21	6.72	0.02	0.00	0.18	6.85
Dtype	0.29	7.27	0.79	15.40	-0.01	0.00	0.65	14.66
Cwall	0.24	1.84	-0.50	-2.69	-0.02	0.00	-0.66	-4.02
Storeys	0.11	2.52	-0.12	-1.91	0.02	0.00	-0.06	-1.01
Roofing	0.25	5.42	-0.81	-12.66	0.06	0.00	-0.57	-10.74
Exwalls	0.54	15.07	0.59	9.44	-0.03	0.00	0.61	10.56
Tla	1410.39	27.83	1215.19	16.19	2457.44	0.00	1611.79	20.13
Totfix	13837.41	14.32	8254.71	8.17	-451.43	0.00	7118.49	6.57
Topo	0.21	1.91	-0.07	-0.53	-2621.68	0.00	0.65	4.23
Sec	1.26	10.59	0.30	2.20	1.47	0.00	0.36	1.62
Erfext	122.22	13.06	185.00	13.12	7.55	0.00	105.59	7.67
R-Squared		0.85		0.96		1.00		0.96
Iterations		13		7		100		16

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	3268	3268	3268	809	809	809	3268	3268	3268	809	809	809
Mean	349556	352308.1	1.09	358672	357301	1.05	349556	354656	1.06	358671.5	357196	1.05
Median	218200	239437	1.04	222860	236206	1.02	218200	228177	1.04	222860	230000	1.02
Minimum	33000	26575.2	0.25	35958	37134.2	0.47	33000	45127.6	0.89	35958	41917.3	0.55
Lower Quartile	106500	122298.3	0.89	107730	116822	0.86	106500	120405	0.99	107730	120035	0.89
Upper Quartile	425460	422804.5	1.23	420420	407641	1.19	425460	422051	1.12	420420	410133	1.18
Maximum	6858000	4515618	3.25	8280000	8096804	3.34	6858000	6990855	1.63	8280000	8281096	2.85
Range	6825000	4489043	3.00	8244042	8059670	2.87	6825000	6945728	0.75	8244042	8239179	2.30
IQR	318960	300506.2	0.34	312690	290819	0.33	318960	301646	0.12	312690	290098	0.29
Variance	1.70E+11	1.42E+11	0.09	2.86E+11	2.75E+11	0.08	1.70E+11	1.66E+11	0.01	2.86E+11	2.76E+11	0.06
Std. Deviation	412697	377304.1	0.29	534472	524524	0.29	412697	407843	0.09	534471.5	525171	0.24
CoV	1.18063	1.070949	0.27	1.49014	1.46802	0.27	1.18063	1.14997	0.09	1.490142	1.47026	0.23
Ave |Dev. Med|	231371	220796.8	0.22	239960	229182	0.21	231371	224329	0.07	239959.5	230519	0.18
CoD	106.04	92.21	20.92	107.67	97.03	20.43	106.04	98.31	6.89	107.67	100.23	17.32
Weighted Mean			1.01			1.00			1.01			1.00
PRD			107.8			105.7			104.9			105.1

	LOCAL				ORDINARY			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	1.09	99.82	0.47	19.47	1.28	70.87	1.07	34.83
View	0.06	3.72	0.92	25.53	-0.06	-3.32	0.36	10.98
Cond_Qual	0.12	13.29	0.23	5.13	0.09	8.66	0.19	7.13
Dtype	-0.08	-3.65	0.88	12.04	0.09	3.99	0.58	13.35
Cwall	-0.11	-1.63	-0.81	-3.05	-0.09	-1.14	-0.56	-3.45
Storeys	-0.06	-2.83	0.26	2.74	-0.03	-1.21	-0.13	-2.40
Roofing	0.02	0.91	-0.64	-7.53	0.08	2.84	-0.55	-10.56
Exwalls	0.00	0.15	1.33	14.65	0.04	1.81	0.59	10.36
Tla	1997.22	68.15	1206.01	12.45	1832.58	59.28	1522.98	19.95
Totfix	6993.85	11.94	2277.73	1.66	11553.30	16.74	8690.09	8.29
Topo	-0.67	-2.77	-0.30	-1.95	0.42	1.87	0.01	0.05
Sec	0.78	3.94	0.56	4.37	14.46	5.28	0.24	1.07
Erfext	43.73	8.58	252.38	13.02	0.03	0.59	111.26	8.37
R-Squared		0.94		0.94		0.91		0.96
Iterations		28		7		50		15

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	3268	3268	3268	809	809	809	3268	3268	3268	809	809	809
Mean	349556	353607	1.08	358672	362729	1.13	349556	342561	0.98	358672	356191	1.03
Median	218200	228902	1.05	222860	247764	1.10	218200	207606	0.96	222860	229439	1.01
Minimum	33000	52855.8	0.37	35958	50500	0.46	33000	41868.7	0.45	35958	38903.9	0.54
Lower Quartile	106500	119305	0.95	107730	137248	0.92	106500	102410	0.85	107730	115007	0.88
Upper Quartile	425460	427037	1.17	420420	403071	1.30	425460	413986	1.09	420420	413264	1.15
Maximum	6858000	7162336	2.39	8280000	8302560	3.37	6858000	5711414	2.20	8280000	8229043	2.77
Range	6825000	7109480	2.02	8244042	8252060	2.91	6825000	5669545	1.75	8244042	8190140	2.23
IQR	318960	307733	0.22	312690	265823	0.38	318960	311576	0.24	312690	298257	0.27
Variance	1.70E+11	1.58E+11	0.04	2.86E+11	2.65E+11	0.11	1.70E+11	1.60E+11	0.04	2.86E+11	2.77E+11	0.05
Std. Deviation	412697	396899	0.19	534472	514460	0.33	412697	400250	0.20	534472	526094	0.23
CoV	1.18063	1.12243	0.18	1.49014	1.4183	0.29	1.18063	1.16841	0.20	1.49014	1.477	0.22
Ave Dev. Med	231371	223139	0.14	239960	220495	0.25	231371	229324	0.15	239960	233366	0.17
CoD	106.04	97.48	13.36	107.67	88.99	22.45	106.04	110.46	15.42	107.67	101.71	16.87
Weighted Mean			1.01			1.01			0.98			0.99
PRD			106.4			111.8			100.4			103.4

ATM Model

	GLOBAL				IDW			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	0.61	10.36	0.91	10.20	1.14	330.38	0.92	11.21
Str	0.76	4.05	1.00	3.68	-0.07	-4.29	0.70	2.55
View	0.27	4.95	0.73	7.76	0.00	-0.03	0.30	2.92
Cond_Qual	0.14	3.94	0.37	4.69	0.02	7.52	0.40	5.69
Exwalls	0.48	5.89	0.61	3.65	-0.02	-4.30	0.68	4.26
Tla	1815.94	12.87	1089.10	4.97	2534.38	238.76	1483.11	6.14
Totfix	17144.35	5.77	8980.04	2.76	-507.57	-2.31	5618.03	1.55
Sec	1.00	4.00	-0.25	-0.64	-0.96	-2.19	-0.13	-0.29
Topo	0.08	0.33	0.02	0.06	-80.82	0.00	0.74	3.04
Erfext	162.20	6.40	279.13	5.30	15.07	9.65	205.92	3.88
R-Squared		0.69		0.91		1.00		0.92
Iterations		13		15		50		16

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	604	604	604	147	147	147	604	604	604	147	147	147
Mean	942548.1	935065.2	1.04	994945.4	965640.6	0.94	942548.1	942860.7	1.00	994945.4	970839.4	0.96
Median	794750	828721.9	1.01	786080	726036.6	0.91	794750	803441.8	1.00	786080	694877.2	0.93
Minimum	180000	197661	0.33	248940	155797	0.47	180000	183128.5	0.88	248940	175638	0.51
Lower Quartile	568229.6	573536.4	0.85	536310	437683.7	0.73	568229.6	569215.9	0.99	536310	488505.7	0.75
Upper Quartile	1135050	1121063	1.18	1128600	1259165	1.13	1135050	1133223	1.01	1128600	1188037	1.13
Maximum	6858000	4767016	3.20	8280000	7910081	1.71	6858000	6890671	1.17	8280000	8164248	1.63
Range	6678000	4569355	2.87	8031060	7754284	1.24	6678000	6707543	0.28	8031060	7988610	1.12
IQR	566820.4	547526.8	0.33	592290	821480.8	0.40	566820.4	564006.8	0.03	592290	699531.5	0.37
Variance	3.88E+11	2.84E+11	0.08	9.71E+11	9.45E+11	0.08	3.88E+11	3.87E+11	0.00	9.71E+11	9.41E+11	0.06
Std. Deviation	623281.9	532891.4	0.29	985515.9	972255.8	0.28	623281.9	622028.3	0.03	985515.9	970111.7	0.25
CoV	0.66	0.57	0.28	0.99	1.01	0.29	0.66	0.66	0.03	0.99	1.00	0.26
Ave |Dev. Med	380912.6	364428.8	0.21	450936.3	523380.7	0.22	380912.6	379362.8	0.02	450936.3	507880.3	0.20
CoD	47.93	43.97	20.83	57.37	72.09	24.48	47.93	47.22	1.87	57.37	73.09	21.51
Weighted Mean			0.99			0.97			1.00			0.98
PRD			105.0			96.9			100.2			97.9

	LOCAL				ORDINARY			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**
LocMult	1.00	35.46	0.43	6.60	1.32	25.96	1.11	12.14
Str	0.20	1.68	0.89	2.57	0.38	2.63	0.62	2.37
View	0.07	1.97	0.79	7.46	-0.10	-2.17	0.33	3.41
Cond_Qual	0.10	5.21	0.55	4.91	0.11	5.00	0.35	5.42
Exwalls	0.04	0.83	1.27	6.17	-0.07	-1.22	0.66	4.34
Tla	2041.20	28.14	1210.06	4.16	1757.65	21.91	1341.16	6.27
Totfix	9823.28	6.27	3478.06	0.77	13306.24	7.69	8210.60	2.51
Sec	0.50	1.26	-0.30	-0.95	12.91	2.82	-0.17	-0.34
Topo	-0.58	-1.24	-0.05	-0.20	0.16	0.33	0.23	0.70
Erfext	57.74	4.30	491.76	6.04	0.08	0.38	189.38	3.93
R-Squared		0.88		0.89		0.82		0.93
Iterations		20		19		50		17

	Sale			Sale			Sale			Sale		
	Actual	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**
No of Cases	604	604	604	147	147	147	604	604	604	147	147	147
Mean	942548.1	937611.7	1.02	994945.4	964553.1	0.95	942548.1	933442.8	1.01	994945.4	972220.3	0.96
Median	794750	814202.1	1.00	786080	728010.1	0.89	794750	792836.8	0.98	786080	707718.1	0.94
Minimum	180000	206287.9	0.39	248940	162863.2	0.44	180000	217797.2	0.51	248940	172916.5	0.52
Lower Quartile	568229.6	566440	0.89	536310	433568	0.71	568229.6	544474.1	0.86	536310	489405	0.77
Upper Quartile	1135050	1131476	1.11	1128600	1222937	1.13	1135050	1137529	1.14	1128600	1169209	1.11
Maximum	6858000	7154458	2.35	8280000	7687929	1.86	6858000	5819610	2.19	8280000	8110310	1.62
Range	6678000	6948171	1.95	8031060	7525066	1.42	6678000	5601813	1.69	8031060	7937393	1.10
IQR	566820.4	565036.3	0.22	592290	789369	0.42	566820.4	593054.5	0.28	592290	679804.3	0.34
Variance	3.88E+11	3.51E+11	0.04	9.71E+11	9.2E+11	0.09	3.9E+11	3.4E+11	0.05	9.7E+11	9.44E+11	0.06
Std. Deviation	623281.9	592360.3	0.19	985515.9	959922.2	0.30	623281.9	578552.3	0.23	985515.9	971740.8	0.23
CoV	0.66	0.63	0.19	0.99	1.00	0.31	0.66	0.62	0.22	0.99	1.00	0.25
Ave Dev. Med	380912.6	370914.8	0.14	450936.3	515606.1	0.24	380912.6	387120.6	0.17	450936.3	501780.6	0.19
CoD	47.9286	45.55562	13.57	57.36519	70.82403	26.92	47.9286	48.82727	17.40	57.36519	70.90118	20.35
Weighted Mean			0.99			0.97			0.99			0.98
PRD			102.5			98.1			102.4			98.1

Not Around Table Mountain

	GLOBAL				IDW			
	Modelling Sample		Test Sample		Modelling Sample		Testing Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	1.05	44.28	1.01	22.86	1.37	677.34	1.11	28.14
View	0.01	0.31	0.23	3.92	-0.03	-11.26	0.28	5.53
Cond_Qual	0.20	8.41	0.38	8.35	0.00	1.64	0.27	7.13
Dtype	0.31	5.93	-0.04	-0.42	-0.02	-4.17	0.05	0.60
Exwalls	0.12	2.05	0.59	5.55	0.03	7.15	0.25	2.70
Tla	1352.35	35.33	1417.05	18.01	2430.80	667.86	1704.40	25.25
Totfix	6891.20	11.54	6039.72	5.23	-194.32	-3.03	4247.23	4.06
Str	1.39	5.47	1.07	1.91	0.11	0.03	3.05	3.69
Sec	0.73	9.46	0.22	1.36	-6.72	-0.59	0.10	0.50
Erfext	153.33	18.64	149.78	9.81	1.32	2.03	110.97	8.53
Gard	1.27	2.19	0.05	0.04	-1.25	-0.73	24.98	0.06
Pool	691.69	5.37	466.46	1.91	17.82	1.54	-0.41	-0.01
Garcp	340.42	3.24	331.98	1.65	5.96	0.75	-1.19	-0.01
R-Squared		0.88		0.88		1.00		0.91
Iterations		19		50		38		50

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	2664	2664	2664	662	662	662	2664	2664	2664	662	662	662
Mean	215108.4	213923.4	1.04	217384.1	216526.4	1.04	215108.4	215702	1.01	217384	215610.9	1.03
Median	168232.5	175648	0.99	173062.5	177341.3	1.00	168232.5	168731	1.00	173063	167726.9	1.00
Minimum	33000	21992.04	0.31	35958	31045.35	0.50	33000	38915.9	0.85	35958	51414.79	0.63
Lower Quartile	94945	99165.41	0.86	96050	100775.9	0.88	94945	95812.1	0.99	96050	101217.3	0.88
Upper Quartile	289300	287540.7	1.16	284040	289061.9	1.16	289300	286960	1.02	284040	282950.3	1.13
Maximum	1196800	1027554	2.56	1233750	1294471	3.13	1196800	1230358	1.53	1233750	1141159	2.63
Range	1163800	1005562	2.25	1197792	1263426	2.64	1163800	1191442	0.68	1197792	1089744	2.01
IQR	194355	188375.3	0.29	187990	188286	0.29	194355	191148	0.03	187990	181733	0.25
Variance	2.31E+10	2.08E+10	0.06	2.46E+10	2.21E+10	0.06	2.31E+10	2.29E+10	0.00	2.46E+10	2.31E+10	0.04
Std. Deviation	152131.7	144167.4	0.25	156960.4	148571.9	0.25	152131.7	151193	0.04	156960	151937.4	0.21
CoV	0.71	0.67	0.24	0.72	0.69	0.25	0.71	0.70	0.04	0.72	0.70	0.20
Ave |Dev. Med	113759	109669.6	0.19	116143.2	111932.4	0.18	113759	112330	0.02	116143	111305.9	0.16
CoD	67.62	62.44	18.87	67.11	63.12	18.13	67.62	66.57	2.37	67.11	66.36	15.53
Weighted Mean			0.99			1.00			1.00			0.99
PRD			104.1			104.0			100.9			103.4

	LOCAL				ORDINARY			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**
LocMult	1.38	96.43	1.05	26.58	1.34	75.98	1.16	28.67
View	-0.05	-2.97	0.27	5.19	0.01	0.58	0.24	4.80
Cond_Qual	0.05	4.04	0.26	6.66	0.09	6.17	0.28	7.38
Dtype	0.03	1.05	0.02	0.19	0.12	3.59	-0.02	-0.19
Exwalls	0.09	3.06	0.37	4.00	0.09	2.47	0.25	2.78
Tla	2005.91	82.92	1715.56	24.37	1822.39	63.59	1692.20	25.37
Totfix	2049.76	5.40	3502.95	3.19	3369.09	7.58	3964.69	3.85
Str	1.13	3.38	3.05	3.81	1.33	4.07	2.74	3.59
Sec	0.50	5.03	0.26	1.44	0.55	5.78	0.22	1.27
Erfext	67.84	14.49	115.40	8.50	83.72	14.80	114.50	8.94
Gard	3.94	1.86	24.25	0.06	-0.29	-0.35	22.55	0.09
Pool	168.95	2.34	-0.96	-0.01	221.42	2.66	-1.64	-0.01
Garcp	-91.78	-1.80	-1.25	-0.01	145.40	2.10	-2.07	-0.01
R-Squared		0.95		0.90		0.94		0.91
Iterations		14		50		47		50

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	**Predicted**		**Actual**	**Predicted**		**Actual**	**Predicted**		**Actual**	**Predicted**	
No of Cases	2664	2664	2664	662	662	662	2664	2664	2664	662	662	662
Mean	215108	212549.2	1.00	217384	216030.3	1.03	215108.4	211866.2	0.99	217384	215370	1.02
Median	168233	165801.4	0.98	173063	171767.4	1.01	168232.5	165459.5	0.97	173063	167910	0.99
Minimum	33000	42937.59	0.58	35958	51823.59	0.45	33000	43681.66	0.53	35958	52487.7	0.58
Lower Quartile	94945	94188.35	0.90	96050	102331.5	0.88	94945	93448.98	0.88	96050	97667.3	0.88
Upper Quartile	289300	279879.2	1.07	284040	285206.4	1.15	289300	280107.4	1.08	284040	279896	1.13
Maximum	1196800	1228908	2.34	1233750	1157509	2.70	1196800	1066455	1.83	1233750	1164034	2.62
Range	1163800	1185971	1.76	1197792	1105686	2.24	1163800	1022773	1.29	1197792	1111547	2.04
IQR	194355	185690.8	0.16	187990	182874.9	0.27	194355	186658.5	0.20	187990	182229	0.25
Variance	2.31E+10	2.32E+10	0.02	2.46E+10	2.27E+10	0.05	2.31E+10	2.30E+10	0.03	2.46E+10	2.32E+10	0.04
Std. Deviation	152132	152261.3	0.15	156960	150800.9	0.22	152131.7	151796.5	0.17	156960	152479	0.21
CoV	0.71	0.72	0.15	0.72	0.70	0.21	0.71	0.72	0.17	0.72	0.71	0.20
Ave |Dev. Med	113759	112991.2	0.11	116143	110690.6	0.16	113759	113021.4	0.13	116143	111586	0.15
CoD	67.62	68.15	10.91	67.11	64.44	16.10	67.62	68.31	13.05	67.11	66.46	15.54
Weighted Mean			0.99			0.99			0.98			0.99
PRD			100.8			103.9			101.0			103.2

City of Cape Town Model

	GLOBAL				IDW			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	1.36	32.86	1.82	27.46	0.41	7.81	1.35	20.43
View	0.27	10.77	0.58	16.01	0.46	13.53	0.56	12.92
Cond	0.13	8.20	0.20	7.38	0.14	5.72	0.21	7.23
Dtype	0.10	2.89	0.80	17.39	0.15	3.41	0.68	13.83
Cwall	0.30	2.48	-0.14	-0.78	0.46	3.19	-0.41	-2.10
Storeys	0.08	2.19	-0.32	-5.70	0.08	1.72	-0.31	-4.97
Roofing	0.19	4.67	-0.78	-13.57	0.14	2.78	-0.93	-14.08
Exwalls	0.64	19.04	0.58	9.87	0.81	18.53	0.62	10.22
Tla	1600.55	33.74	1292.03	16.88	1794.30	30.88	1662.32	30.73
Totfix	13814.95	14.29	12505.23	12.34	14775.61	11.68	8970.86	9.92
Topo	0.56	3.06	0.56	2.01	2.01	8.49	12.42	0.97
Sec	2.87	8.94	-0.49	-1.31	1.93	7.04	2.52	2.29
Erfext	30.81	5.17	73.55	5.49	38.09	5.41	0.04	0.11
R-Squared		0.84		0.95		0.79		0.94
Iterations		30		19		13		50

	Sales		Ratio	Sales		Ratio	Sales		Ratio	Sales		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	3268	3268	3268	809	809	809	3268	3268	3268	809	809	809
Mean	349556	345516.5	1.01	358672	346925.4	0.94	349556	355309	1.18	358672	346017.6	0.98
Median	218200	235651	0.97	222860	230180.2	0.92	218200	253970.9	1.10	222860	224087.7	0.94
Minimum	33000	16310.58	0.22	35958	17034.44	0.24	33000	40718.95	0.25	35958	25199.47	0.30
Lower Quartile	106500	101990	0.81	107730	87028.22	0.73	106500	144716.6	0.88	107730	108013.6	0.75
Upper Quartile	425460	424066.4	1.16	420420	404265.6	1.11	425460	426390.3	1.39	420420	398410.1	1.14
Maximum	6858000	4710633	2.98	8280000	8360681	3.37	6858000	5219959	4.68	8280000	8380814	2.74
Range	6825000	4694323	2.75	8244042	8343647	3.13	6825000	5179240	4.43	8244042	8355615	2.44
IQR	318960	322076.5	0.35	312690	317237.4	0.39	318960	281673.7	0.51	312690	290396.5	0.38
Variance	1.70E+11	1.46E+11	0.09	2.86E+11	2.80E+11	0.11	1.70E+11	1.30E+11	0.18	2.86E+11	2.76E+11	0.10
Std. Deviation	412697	381999.5	0.31	534472	529285.1	0.32	412697	360946.6	0.43	534472	525569.8	0.31
CoV	1.18	1.11	0.30	1.49	1.53	0.34	1.18	1.02	0.36	1.49	1.52	0.32
Ave |Dev.Med	231371	228533.8	0.23	239960	240665.2	0.24	231371	208645.9	0.32	239960	232439.6	0.23
CoD	106.04	96.98	23.31	107.67	104.56	26.11	106.04	82.15	28.90	107.67	103.73	24.97
Weighted Mean			0.99			0.97			1.02			0.96
PRD			102.5			97.3			116.2			101.2

	LOCAL				ORDINARY			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)	Parameter Estimate	T-Value (Ho: Bi=0)
LocMult	0.75	21.00	0.53	10.65	1.31	29.60	1.48	20.30
View	0.41	14.94	0.75	16.12	0.29	11.38	0.54	12.92
Cond	0.15	8.06	0.37	9.26	0.08	5.31	0.24	7.32
Dtype	0.09	2.16	0.80	12.12	0.12	3.71	0.74	13.98
Cwall	0.11	0.84	-0.15	-0.61	-0.09	-0.78	-0.57	-2.77
Storeys	0.13	3.17	-0.15	-1.84	-0.13	-3.40	-0.33	-4.98
Roofing	0.07	1.56	-0.86	-10.44	0.07	1.83	-1.04	-14.98
Exwalls	0.73	18.83	0.95	11.85	0.58	16.78	0.65	9.62
Tla	1583.64	30.29	1387.28	12.08	1679.38	39.31	1364.93	16.07
Totfix	14953.79	13.90	13539.14	9.91	14221.38	14.91	10057.04	9.06
Topo	1.59	7.16	2.03	7.29	1.54	4.83	2.03	6.37
Sec	0.69	3.38	0.55	1.54	16.09	3.43	-0.17	-0.43
Erfext	59.51	6.97	75.12	3.70	0.01	0.32	50.62	3.52
R-Squared		0.81		0.92		0.83		0.94
Iterations		22		20		50		13

	Sales		Ratio	Sales		Ratio	Sales		Ratio	Sales		Ratio
	Actual	Predicted		Actual	Predicted		Actual	Predicted		Actual	Predicted	
No of Cases	3268	3268	3268	809	809	809	3268	3268	3268	809	809	809
Mean	349555.6	350930.3	1.12	358672	356251.5	1.10	349556	342462.3	1.01	358672	343983.4	0.95
Median	218200	245949.8	1.05	222860	241274.9	1.05	218200	229916.5	0.96	222860	220211	0.92
Minimum	33000	37262.35	0.27	35958	33763	0.42	33000	27028.53	0.29	35958	22754.98	0.30
Lower Quartile	106500	130698.1	0.86	107730	136914.7	0.85	106500	106391.2	0.80	107730	99538.69	0.75
Upper Quartile	425460	419276.4	1.30	420420	397284.1	1.27	425460	419474.7	1.17	420420	399441	1.11
Maximum	6858000	5020953	3.82	8280000	8541224	3.29	6858000	5061897	3.07	8280000	8472913	2.70
Range	6825000	4983691	3.55	8244042	8507461	2.87	6825000	5034868	2.77	8244042	8450158	2.40
IQR	318960	288578.3	0.44	312690	260369.4	0.42	318960	313083.5	0.37	312690	299902.3	0.37
Variance	1.70E+11	1.38E+11	0.14	2.86E+11	2.64E+11	0.13	1.70E+11	1.46E+11	0.10	2.86E+11	2.79E+11	0.09
Std. Deviation	412697	370836.1	0.37	534472	514034	0.37	412697	382715.3	0.32	534472	527864.8	0.30
CoV	1.18	1.06	0.33	1.49	1.44	0.33	1.18	1.12	0.32	1.49	1.53	0.32
Ave |Dev.Med	231370.9	214117.1	0.27	239960	219169.5	0.27	231371	224834	0.24	239960	235001.5	0.23
CoD	106.04	87.06	25.94	107.67	90.84	25.90	106.04	97.79	24.74	107.67	106.72	24.93
Weighted Mean			1.00			0.99			0.98			0.96
PRD			111.4			110.7			103.3			99.0

Around Table Mountain

	GLOBAL				IDW			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**
LocMult	1.03	8.16	1.53	6.04	-0.17	-1.44	0.82	3.12
Str	0.70	3.57	0.99	3.10	0.47	2.30	1.28	3.69
View	0.27	4.91	0.70	6.11	0.49	7.18	0.62	4.85
Cond_Qual	0.13	3.81	0.37	4.38	0.23	4.33	0.38	3.63
Exwalls	0.61	7.88	0.74	3.94	0.86	9.72	1.07	5.13
Tla	1891.16	13.33	1094.94	4.42	2395.70	12.73	1413.82	4.34
Totfix	16906.74	5.70	12345.79	3.41	24976.43	5.89	10640.16	2.18
Sec	1.53	3.73	-0.80	-1.20	1.51	3.99	-0.37	-0.65
Topo	0.17	0.50	0.39	0.91	0.80	2.47	1.35	4.05
Erfext	87.73	3.91	188.78	3.26	130.91	3.90	208.27	2.55
R-Squared		0.68		0.88		0.64		0.86
Iterations		29		17		7		15

	Sale		Ratio	Sale		Ratio	Sale		Ratio	Sale		Ratio
	Actual	**Predicted**		**Actual**	**Predicted**		**Actual**	**Predicted**		**Actual**	**Predicted**	
No of Cases	604	604	604	147	147	147	604	604	604	147	147	147
Mean	942548.1	931581.3	1.04	994945	966962	0.95	942548	938470.6	1.07	994945	949111.9	0.93
Median	794750	817237.8	1.00	786080	728051	0.94	794750	825014.9	1.04	786080	689401.1	0.87
Minimum	180000	203152.2	0.29	248940	142076	0.43	180000	209773.3	0.30	248940	126200.4	0.38
Lower Quartile	568229.6	583010.2	0.85	536310	461139	0.74	568230	605757.8	0.86	536310	420549.4	0.68
Upper Quartile	1135050	1119764	1.18	1128600	1259356	1.16	1135050	1118199	1.23	1128600	1169258	1.16
Maximum	6858000	4996040	2.59	8280000	8156114	1.86	6858000	5034846	2.97	8280000	8307497	1.94
Range	6678000	4792887	2.29	8031060	8014037	1.43	6678000	4825073	2.67	8031060	8181297	1.57
IQR	566820.4	536753.4	0.34	592290	798217	0.43	566820	512441.3	0.36	592290	748708.3	0.47
Variance	3.88E+11	2.84E+11	0.08	9.71E+11	9.12E+11	0.08	3.88E+11	2.58E+11	0.09	9.71E+11	9.27E+11	0.10
Std. Deviation	623281.9	533307.9	0.29	985516	955191	0.29	623282	507935.9	0.30	985516	962987.8	0.31
CoV	0.66	0.57	0.28	0.99	0.99	0.31	0.66	0.54	0.28	0.99	1.01	0.33
Ave |Dev. Med	380912.6	361287.5	0.21	450936	518811	0.24	380913	342427.6	0.23	450936	517771.2	0.25
CoD	47.93	44.21	21.32	57.37	71.26	25.23	47.93	41.51	21.70	57.37	75.10	28.35
Weighted Mean			0.99			0.97			1.00			0.95
PRD			105.2			97.8			107.1			97.4

	LOCAL				ORDINARY			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable Name	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**	**Parameter Estimate**	**T-Value (Ho: Bi=0)**
LocMult	0.43	4.99	0.18	1.10	1.02	7.93	0.99	3.25
Str	0.34	1.71	1.46	3.99	0.42	2.20	1.29	3.74
View	0.37	6.56	0.72	5.70	0.27	4.71	0.62	4.91
Cond_Qual	0.14	3.85	0.52	4.44	0.08	2.46	0.38	3.78
Exwalls	0.70	8.55	1.27	5.90	0.55	7.31	1.02	4.85
Tla	2052.17	13.04	1426.67	3.68	1972.28	14.09	1315.12	4.32
Totfix	22119.69	6.50	12489.03	2.25	20550.08	7.19	9887.05	2.17
Sec	0.57	1.60	-0.30	-0.62	12.81	2.17	-0.55	-0.95
Topo	0.66	1.83	1.09	3.90	1.34	2.00	1.14	3.58
Erfext	123.44	4.37	339.88	3.12	0.08	0.29	209.69	2.70
R-Squared		0.66		0.85		0.67		0.86
Iterations		25		16		48		14

	Sale			Sale			Sale			Sale		
	Actual	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**	**Actual**	**Predicted**	**Ratio**
No of Cases	604	604	604	147	147	147	604	604	604	147	147	147
Mean	942548.1	923561.9	1.03	994945.4	956377.8	0.95	942548	920551.7	1.02	994945.4	949375.3	0.93
Median	794750	810644.7	0.99	786080	701190.4	0.89	794750	801148.1	0.98	786080	703512.3	0.87
Minimum	180000	182496.6	0.32	248940	163003.9	0.41	180000	191515.5	0.32	248940	152158.6	0.36
Lower Quartile	568229.6	570368.3	0.83	536310	440775.5	0.72	568230	545606.9	0.82	536310	445183.5	0.70
Upper Quartile	1135050	1117894	1.20	1128600	1145178	1.17	1135050	1129144	1.17	1128600	1191434	1.13
Maximum	6858000	4645902	3.08	8280000	8309746	1.98	6858000	5075859	2.81	8280000	8284742	2.00
Range	6678000	4463405	2.77	8031060	8146742	1.56	6678000	4884343	2.49	8031060	8132584	1.65
IQR	566820.4	547526	0.38	592290	704402.3	0.45	566820	583537.3	0.36	592290	746250.2	0.43
Variance	3.88E+11	2.91E+11	0.09	9.71E+11	9.06E+11	0.10	3.88E+11	3.01E+11	0.09	9.71E+11	9.28E+11	0.09
Std. Deviation	623281.9	539852.8	0.31	985515.9	951830.4	0.31	623282	548557.6	0.30	985515.9	963192.2	0.30
CoV	0.66	0.58	0.30	0.99	1.00	0.33	0.66	0.60	0.29	0.99	1.01	0.33
Ave |Dev. Med	380912.6	363892.6	0.22	450936.3	510226.4	0.25	380913	378188.3	0.22	450936.3	518498.9	0.25
CoD	47.93	44.89	22.45	57.37	72.77	27.72	47.93	47.21	22.85	57.37	73.70	28.26
Weighted Mean			0.98			0.96			0.98			0.95
PRD			105.2			98.8			104.2			97.3

Not Around Table Mountain

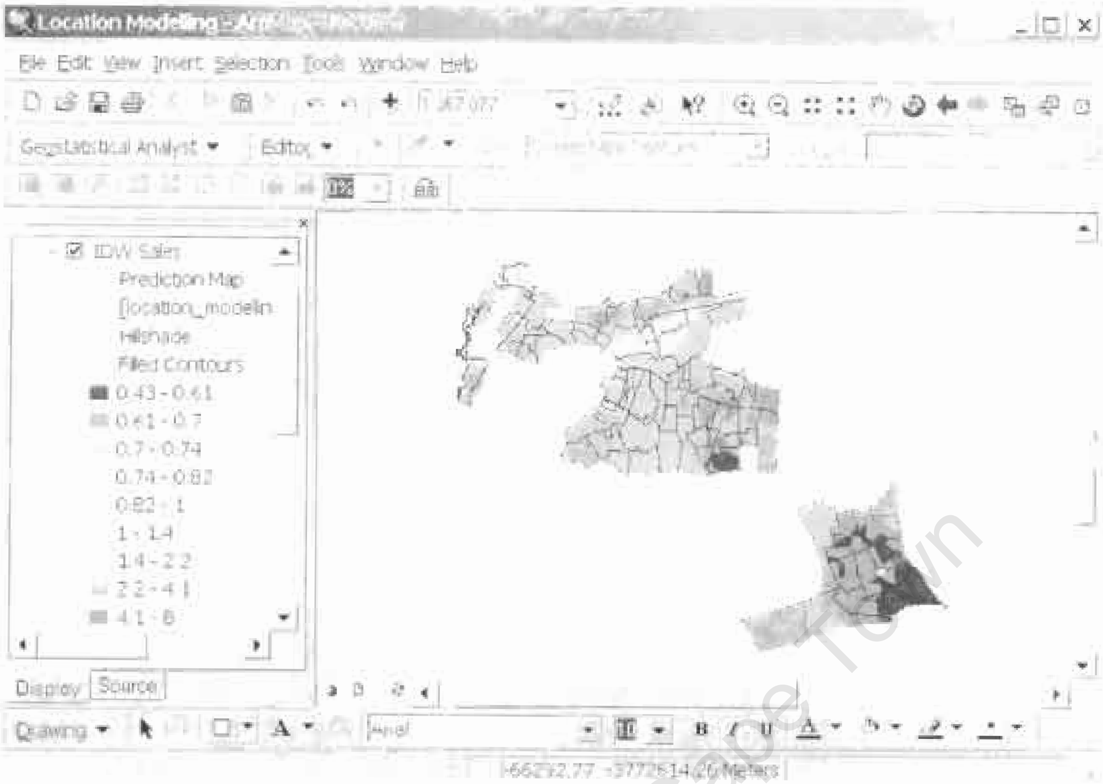
	GLOBAL INTERPOLATOR				IDW INTERPOLATOR			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable	Parameter	T-Value	Parameter	T-Value	Parameter	T-Value	Parameter	T-Value
Name	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)
LocMult	1.53	39.61	1.46	19.68	0.61	12.80	1.15	18.38
View	0.21	7.79	0.38	6.42	0.29	7.66	0.44	7.36
Cond_Qual	0.13	5.68	0.34	7.49	-0.14	-3.69	0.21	4.53
Dtype	0.33	6.47	-0.11	-1.10	0.40	5.80	0.01	0.10
Exwalls	0.01	0.18	0.54	5.30	0.30	3.59	0.46	4.53
Tla	1364.22	37.17	1455.75	19.30	1270.05	28.18	1512.94	20.44
Totfix	6822.29	11.81	5449.74	4.78	5249.72	7.01	6139.26	5.39
Str	1.46	4.55	1.28	1.86	1.49	4.64	2.81	2.57
Sec	0.89	9.12	0.15	0.78	0.48	4.14	-0.33	-1.11
Erfext	112.07	14.51	119.91	8.58	151.17	15.51	91.04	6.34
Gard	0.83	0.73	-15.24	-0.09	-0.59	-0.56	-27.70	-0.01
Pool	453.25	3.85	-0.66	-0.01	510.52	3.30	0.00	0.00
Garcp	55.90	0.57	-0.03	-0.01	59.38	0.61	0.00	0.00
R-Squared		0.87		0.87		0.79		0.86
Iterations		30		22		50		50

	Sale			Sale			Sale			Sale		
	Actual	Predicted	Ratio	Actual	Predicted	Ratio	Actual	Predicted	Ratio	Actual	Predicted	Ratio
No of Cases	2664	2664	2664	662	662	662	2664	2664	2664	662	662	662
Mean	215108.4	211018	1.00	217384.1	214059.3	1.01	215108	215617.7	1.11	217384.1	214667.1	1.03
Median	168232.5	178250	0.94	173062.5	183281.6	0.97	168233	180759	1.06	173062.5	181509.6	0.99
Minimum	33000	18756.2	0.26	35958	29764.83	0.48	33000	48608.77	0.44	35958	46669.14	0.49
Lower Quartile	94945	86568.5	0.81	96050	90901.44	0.82	94945	114303	0.86	96050	100763.5	0.85
Upper Quartile	289300	292469	1.13	284040	294508.1	1.14	289300	279910.1	1.30	284040	283405.8	1.16
Maximum	1196800	1018884	2.84	1233750	1271597	3.34	1196800	1043126	2.96	1233750	1204427	2.69
Range	1163800	1000128	2.57	1197792	1241832	2.86	1163800	994517.6	2.52	1197792	1157758	2.20
IQR	194355	205900	0.32	187990	203606.7	0.32	194355	165607.1	0.43	187990	182642.3	0.31
Variance	2.31E+10	2.18E+10	0.08	2.46E+10	2.29E+10	0.08	2.31E+10	1.81E+10	0.11	2.46E+10	2.24E+10	0.07
Std. Deviation	152131.7	147765	0.27	156960.4	151221.8	0.28	152132	134535.6	0.34	156960.4	149517.7	0.27
CoV	0.71	0.70	0.27	0.72	0.71	0.28	0.71	0.62	0.30	0.72	0.70	0.27
Ave |Dev. Med	113759	114938	0.20	116143.2	115904	0.20	113759	98931.09	0.26	116143.2	110858.2	0.20
CoD	67.62	64.48	21.21	67.11	63.24	20.78	67.62	54.73	24.33	67.11	61.08	20.58
Weighted Mean			0.98			0.98			1.00			0.99
PRD			101.8			102.2			110.4			104.4

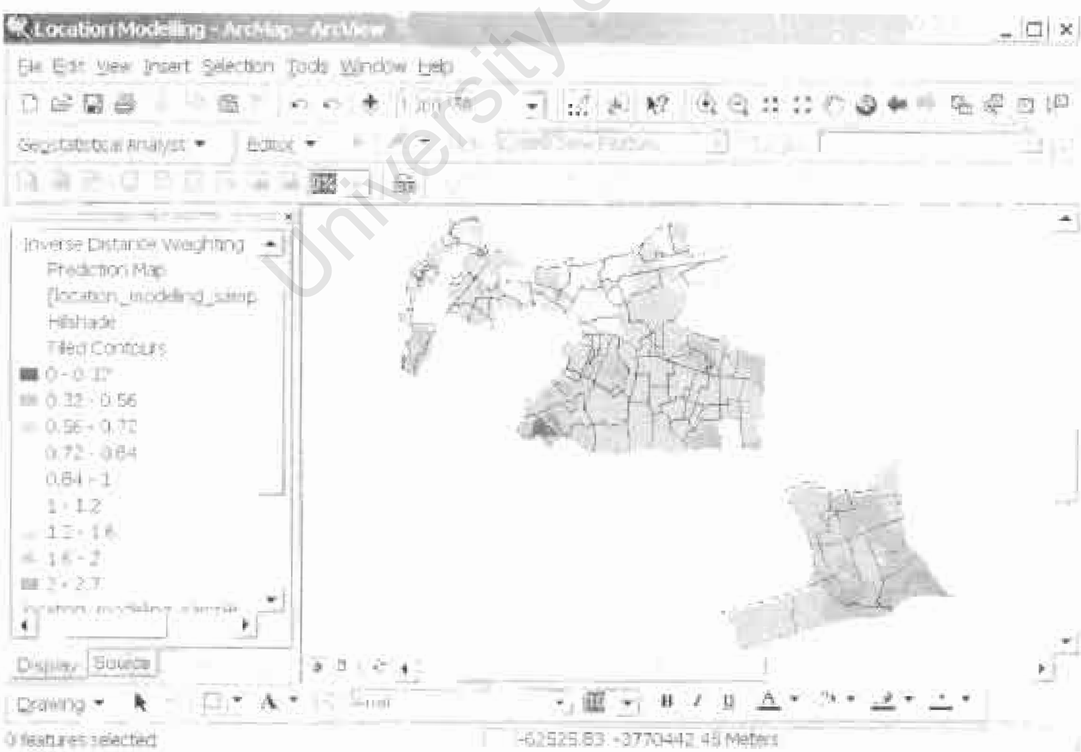
	LOCAL INTERPOLATOR				ORDINARY KRIGING			
	Modelling Sample		Test Sample		Modelling Sample		Test Sample	
Variable	Parameter	T-Value	Parameter	T-Value	Parameter	T-Value	Parameter	T-Value
Name	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)	Estimate	(Ho: Bi=0)
LocMult	1.06	26.86	1.16	18.08	1.21	32.87	1.20	19.34
View	0.20	6.09	0.46	7.65	0.21	7.07	0.42	7.07
Cond_Qual	-0.14	-4.96	0.25	5.26	-0.06	-2.35	0.20	4.20
Dtype	0.16	2.68	-0.06	-0.58	0.19	3.45	0.00	0.03
Exwalls	0.17	2.63	0.47	4.38	0.12	2.04	0.43	4.32
Tla	1411.21	34.12	1453.20	19.77	1453.15	36.94	1495.41	20.24
Totfix	4832.52	7.36	5263.75	4.54	5333.02	8.60	5415.33	4.83
Str	1.84	4.89	2.86	3.17	1.93	4.73	2.69	2.66
Sec	0.38	3.41	-0.06	-0.25	0.50	4.24	-0.23	-0.84
Erfext	117.00	13.57	112.74	7.66	99.66	12.23	95.79	6.68
Gard	-0.18	-0.09	7.13	2.08	0.42	0.19	-2.42	-0.32
Pool	132.25	1.06	3.57	0.05	-0.94	-0.01	93.50	0.36
Garcp	102.56	1.01	-219.43	-1.02	133.50	1.31	2.57	0.03
R-Squared		0.83		0.86		0.85		0.86
Iterations		33		50		50		32

	Sale			Sale			Sale			Sale		
	Actual	Predicted	Ratio	Actual	Predicted	Ratio	Actual	Predicted	Ratio	Actual	Predicted	Ratio
No of Cases	2664	2664	2664	662	662	662	2664	2664	2664	662	662	662
Mean	215108.4	212957.9	1.06	217384.1	215005	1.04	215108	211826	1.03	217384	214329.5	1.03
Median	168232.5	177562.7	1.01	173062.5	184193	0.99	168233	175930	0.98	173063	181075.4	0.99
Minimum	33000	43308.59	0.44	35958	27767.2	0.49	33000	41105.9	0.46	35958	46740.27	0.49
Lower Quartile	94945	102380.4	0.85	96050	100032	0.85	94945	98220.6	0.84	96050	98861.64	0.84
Upper Quartile	289300	279385.2	1.21	284040	285692	1.18	289300	276060	1.17	284040	287830.3	1.16
Maximum	1196800	1203943	2.74	1233750	1151572	2.87	1196800	1110375	2.65	1233750	1219301	2.74
Range	1163800	1160635	2.30	1197792	1123805	2.37	1163800	1069269	2.19	1197792	1172560	2.25
IQR	194355	177004.8	0.35	187990	185660	0.33	194355	177840	0.33	187990	188968.6	0.32
Variance	2.31E+10	2.01E+10	0.09	2.46E+10	2.21E+10	0.08	2.31E+10	2.10E+10	0.08	2.46E+10	2.26E+10	0.07
Std. Deviation	152131.7	141768.7	0.30	156960.4	148732	0.28	152132	144880	0.28	156960	150226.1	0.27
CoV	0.71	0.67	0.28	0.72	0.69	0.27	0.71	0.68	0.27	0.72	0.70	0.26
Ave |Dev. Med	113759	104657.3	0.22	116143.2	109911	0.21	113759	107598	0.21	116143	110837.7	0.20
CoD	67.62	58.94	21.92	67.11	59.67	21.12	67.62	61.16	21.06	67.11	61.21	19.95
Weighted Mean			0.99			0.99			0.98			0.99
PRD			106.7			104.7			104.9			104.2

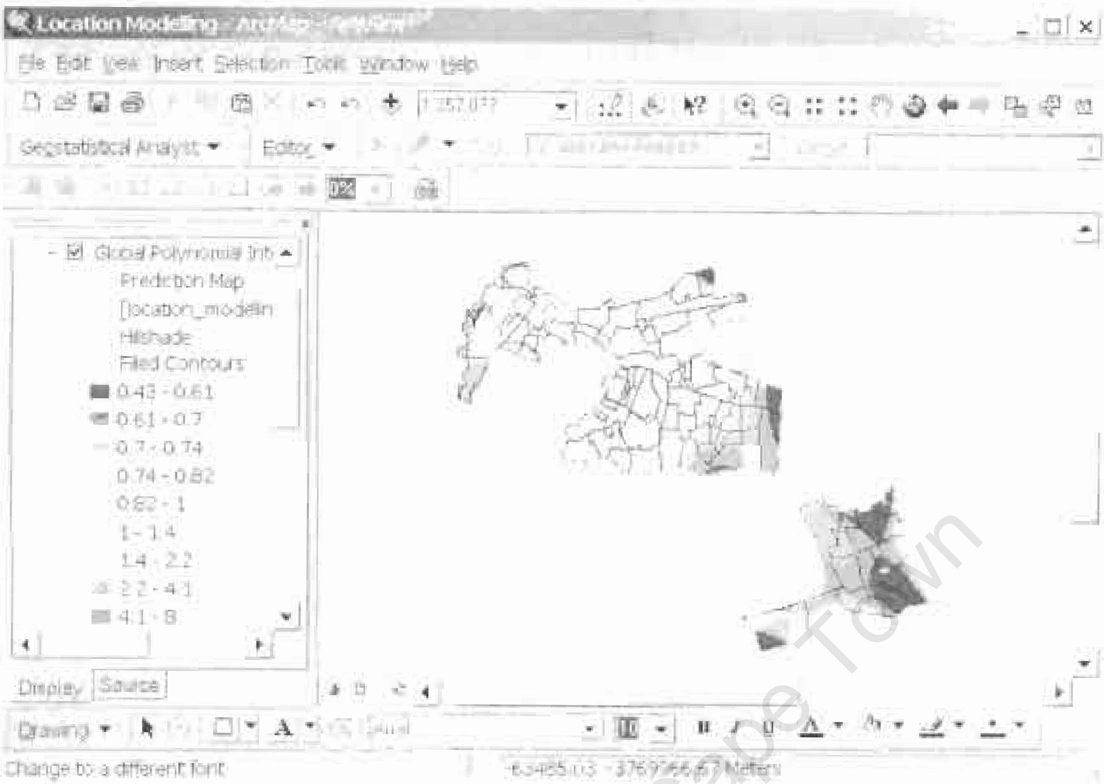
i. Zscores derived from SPSM



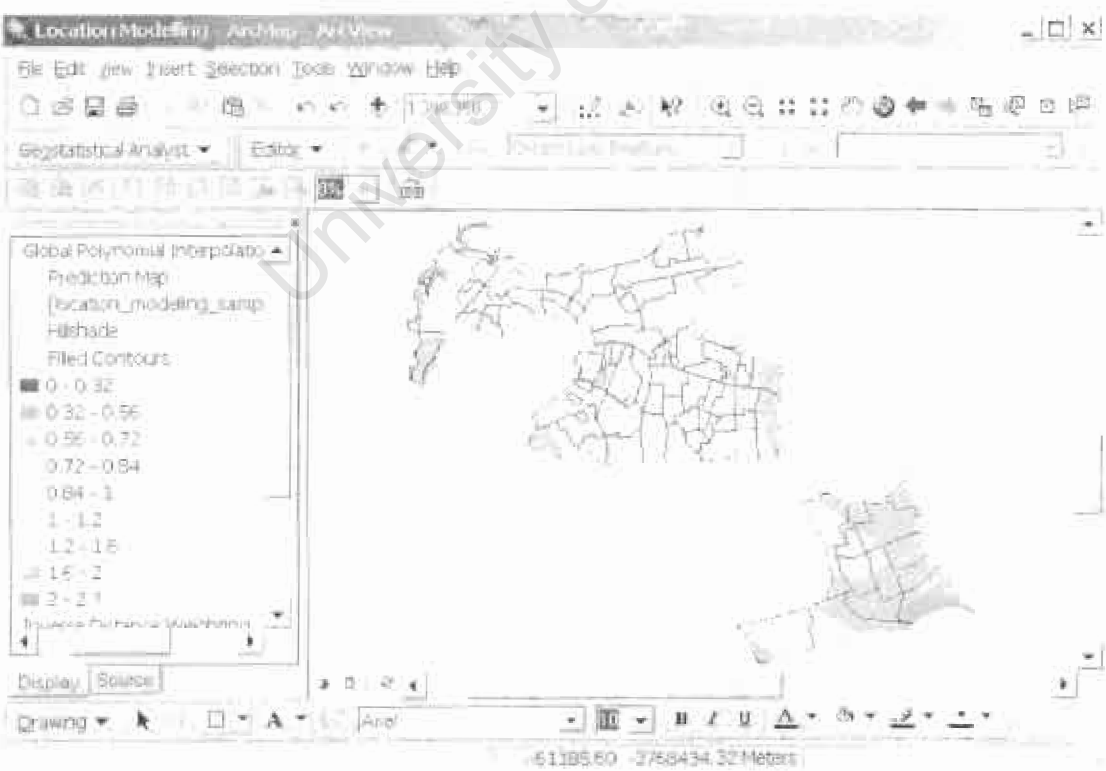
ii. Zscores derived from property characteristics



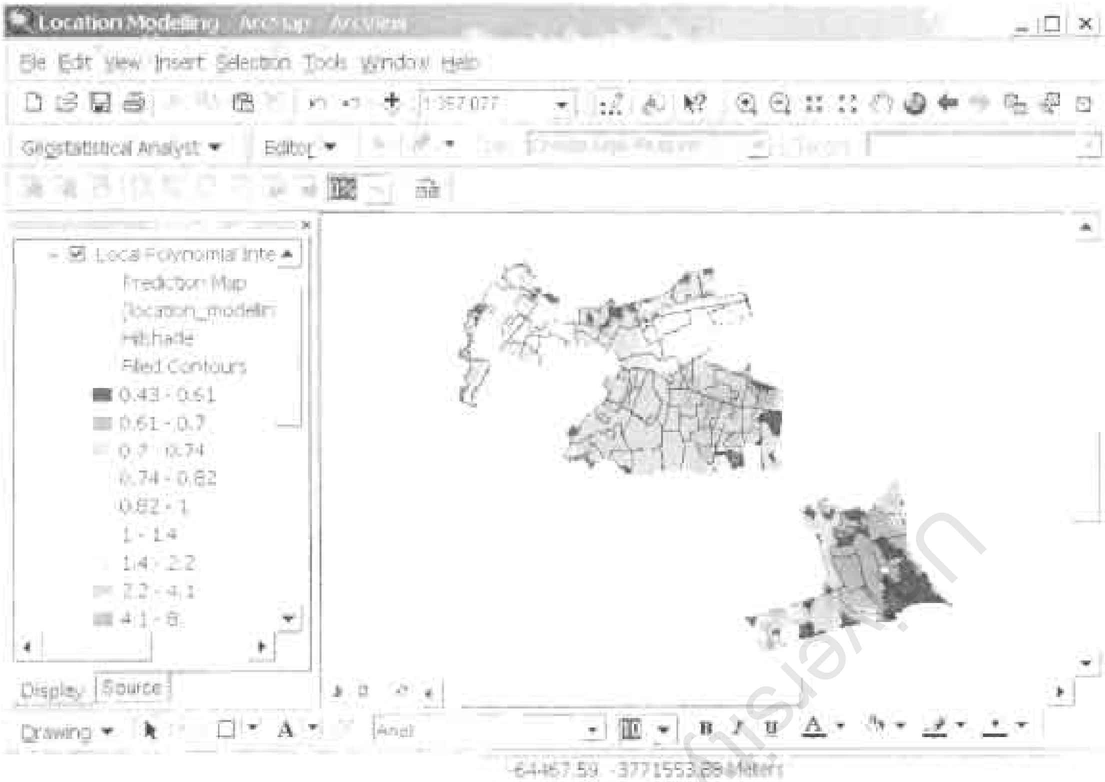
i. Zscores derived from SPSM



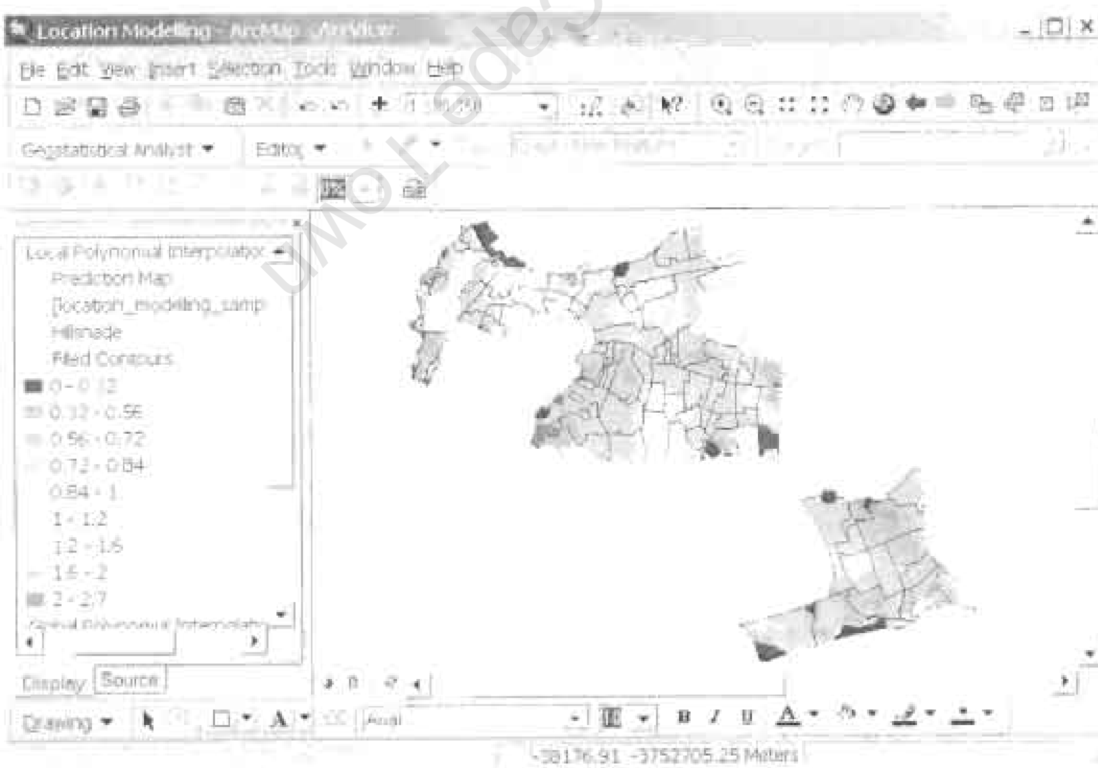
ii. Zscores derived from property characteristics



i. Zscores derived from SPSM

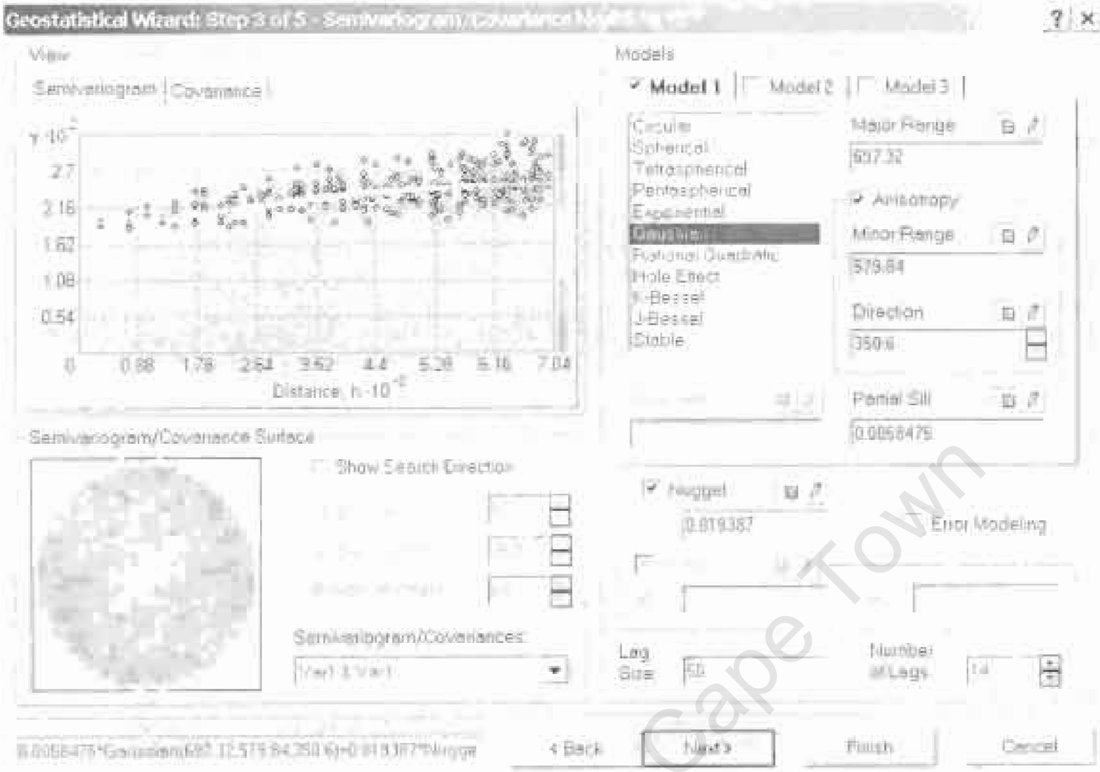


ii. Zscores derived from property characteristics



a. Semivariograms

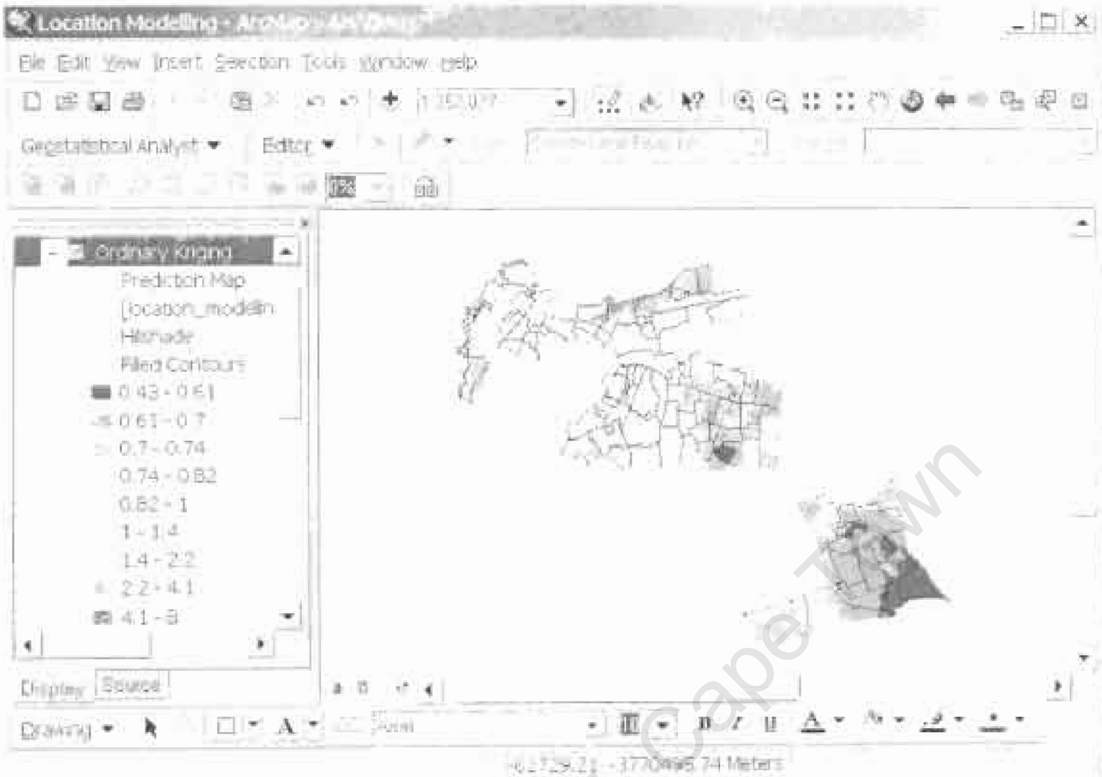
i. Zscores derived from SPSM



ii. Zscores derived from property characteristics

b. Response Surfaces

i. Zscores derived from SPSM



ii. Zscores derived from property characteristics