UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ARTIFICIAL NEURAL NETWORKS MODELLING FOR MASS APPRAISAL OF PROPERTIES

YACIM JOSEPH AWOAMIM

STUDENT NO: 14052190

ARTIFICIAL NEURAL NETWORKS MODELLING FOR MASS APPRAISAL OF PROPERTIES

YACIM JOSEPH AWOAMIM

STUDENT NO: 14052190

A thesis submitted for the degree of

Philosophiae Doctor (PhD) in Real Estate

Department of Construction Economics

Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

Supervisor:

Dr Douw Boshoff

2017

ABSTRACT

This thesis extends the use of artificial neural networks (ANNs) optimisation and training algorithms including the Powell-Beale conjugate gradient (PBCG), scaled conjugate gradient (SCG) and a hybrid system of particle swarm optimisation (PSO) with the traditional back propagation (BP) in mass appraisal as a first attempt. The goal is to verify the comparative performance of ANNs with the traditional hedonic regression and some other modelling techniques including geographically weighted regression (GWR), spatial error model (SEM), spatial lag model (SLM), additive nonparametric regression (ANR), M5P trees and the support vector machines (SVMs). The methodologies are applied to data of 3232 sales transaction of single-family dwellings sold during the period, January 2012 to May 2014 in Cape Town, South Africa. The analysis was done in categories such that the best performing method in each category is selected for a final comparative analysis. The results reveal that semi-log model, SEM, normalised polynomial kernel function support vector machines (NPKSVMs), ANR and the Levenberg-Marquardt trained artificial neural networks (LMANNs) performed best in their respective category. The study also demonstrates the practicability of building hybrid systems in mass appraisal, unfortunately, the hybrid models produces an unexpected results relative to the standalone ANN models. Furthermore, the five best performed models were subjected to three different tests namely, prediction accuracy within the 10 and 20%, model performance and reliability ranking order and lastly explicit explainability ranking order. The final results reveal the LMANNs to outperform the ANR, semi-log, SEM and SVMs in the first two tests, but when the explicit explainability ranking order test which consist of simplicity, consistency, transparency, locational and applicability within the mass appraisal environment was performed, the LMANNs failed the test. The results demonstrate the SEM as the most preferred technique because of its transparency, locational advantage and ease of application within the mass

appraisal environment. Furthermore, it is inferred from the findings that having superior predictive power is imperative, but most importantly is whether the model can practically and effectively be used in mass appraisal of properties. The black box nature of the ANNs inhibits the production of sufficiently transparent estimates that appraisers could use to explain the process when required as a defence before a tribunal or in a formal court.

This thesis contributes to knowledge as follows:

i. Analyse the significance of spatial variation of property prices, with Cape Town, South Africa used as case study;

ii. Build a hybrid system of PSO and BP in mass appraisal;

iii. Improve the training of ANNs in mass appraisal with SCG and PBCG algorithms; and,

iv. Extend the use of GWR, SEM, SLM, SVMs, ANR and log transformation of variables into the South African property market context.

Key words: Artificial neural network, back propagation, hedonic regression, training algorithms, hybrid models, effective predictions.

ACKNOWLEDGEMENT

I would like to specially thank my supervisor, Dr Douw Boshoff for providing good leadership, supervision and ensuring that everything needed for a successful conduct of high level doctoral thesis at University of Pretoria is made available to me. I also wish to thank the head of department, Prof Marthinus Maritz, for providing enabling environment and financial support for the conduct of this research and conference attendance. I am deeply grateful to Mrs Adri Viljoen for helping me to calibrate the Table of contents and list of Tables. I would also like to thank Mrs Alta Lotz, Ms Marveline Molema, Dr Michelle Burger, Mrs Miechaine Buys and many others in the department for making it a wonderful place for me.

I am grateful to my employer Federal Polytechnic, Nasarawa, Nigeria for the permission and support to undertake this study. The efforts of the registrar, Mrs Fatu J. Sabo towards this study is greatly acknowledged. Unfortunately Dr Hassan Sani did not live to see the completion of my doctoral study, his efforts and encouragements are deeply appreciated. The support and encouragement of my head of department, Mall Abdullahi Ahmed is also acknowledged

I acknowledge with thanks Professor M.O. Bello and Associate Professor V.A. Bello of Federal University of Technology, Akure, Nigeria for laying the foundation for a high level research in me during my master of technology degree which culminated into this doctoral thesis. I also want to thank the city valuation office, Cape Town for the property data supplied for this assessment. In particular the efforts of Justine Kahonde, a staff of city valuation office is duly acknowledged for the data and other useful information provided

I would like to thank my parents Mr and Mrs Emmanuel L. Yacim for their encouragement and prayers. I reserved a special place and appreciation to my wife, Mrs Janet Funke Yacim, for her role in keeping the home front while I was away for this study.

My children are also greatly appreciated for their patience and moral support throughout the period of my study.

I would like to appreciate my brethren at Deeper Life Campus Fellowship (DLCF), University of Pretoria for the warmth fellowship we had together. Above all, the Most High God is worthy of my praises. He gave me insights into formulation of the concept used in this study. Your name shall continually be praised throughout all generations as the giver of life, wisdom and knowledge in Jesus name.

Although opinions and findings are those of the author, the financial assistance of National Research Foundation (NRF) and IRE/BS foundation for African Real Estate Research towards this research is hereby acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criteria
ANNs	Artificial Neural Networks
ANR	Additive Nonparametric Regression
AVM	Automated Valuation Models
BP	Back Propagation
CAMA	Computer Assisted Mass Appraisal
CG	Conjugate Gradient
CC	Correlation Coefficients
COD	Coefficient of Dispersion
CAR	Conditional Autoregressive
CS	Cuckoo Search
CSBP	Cuckoo Search Back Propagation
CSLM	Cuckoo Search Levenberg-Marquardt
CVO	City Valuation Office
GA	Genetic Algorithm
GIS	Geographical Information System
GWR	Geographically Weighted Regression
HRM	Hedonic Regression Model
LM	Levenberg-Marquardt
M5P	Decision Tree Regression
MAE	Mean Absolute Error
NNSVM	Neural Network Support Vector Machine
OLS	Ordinary Least Squares
PBCG	Powell-Beale Conjugate Gradient

PRD	Price Related Differential
PSO	Particle Swarm Optimisation
PSOBP	Particle Swarm Optimisation Back Propagation
RMSE	Root Mean Squared Error
RSA	Republic of South Africa
SAR	Simultaneous Autoregressive
SEM	Spatial Error Model
SLM	Spatial Lag Model
SVMs	Support Vector Machines
SCG	Scaled Conjugate Gradient

1. CHAPTER ONE

1.0 Introduction

The artificial neural networks became a model of interest in the field of real estate and valuation for more than two decades now. Pioneering works in the field included Borst (1991); Do & Grudnitski (1992); Tay & Ho (1992); Borst (1995); Worzala, Lenk & Silva (1995) and McCluskey (1996). The model is designed to handle the complex nonlinear relationship that exists in data without the parametric restrictions that are found in statistical techniques. This chapter provides the background leading to the study, as well as the research objectives and methodology, and finally, the thesis structure and contribution to knowledge.

1.1 Background to the Study

The market value of a property is a matter of great interest to local authorities, mortgage institutions, dissolved companies and other market participants, as either of the parties might be disadvantaged should there be an error in the assessment process. Though litigation arising from inconsistent and unreliable estimates by disadvantage parties rarely occurs, it is necessary to guard against its occurrence by ensuring that estimates reflect the market price of properties. Appraisal of a property or properties is a complex procedure due to the different influential factors that constitute the market price(s). While it appears relatively easy to conceptualise the features that will most considerably be associated with a property market price, quantifying their magnitude and contributions is another difficulty. Traditionally, income, cost and market approaches are utilised in the estimation of market values of residential properties. But these methods are increasingly becoming unsustainable for mass appraisal of properties because of valuers' subjectivity, time, cost and insufficient number of comparable properties for assessment. Having realised these limitations particularly as it concerns cost, time and accuracy, various municipalities have introduced computer assisted mass appraisal (CAMA) with the use of hedonic regression models (HRMs). In South Africa, the most successful application is probably the city valuation office (CVO) Cape Town. According to KPMG 2015 report, modernisation of CVO resulted in the reduction of the general assessment cost by R94m (US$7.7m) from 2000 to 2009. Consequently the total revenue from property tax that accrue to the city of Cape Town for 2014 alone was R6 billion (US$ @ 12.21).

The study of Bourassa, Cantoni & Hoesli (2010: 139) reported the widespread use of the HRMs to include price index construction, mass appraisal of properties for taxation, mortgage underwriting and portfolio management. The study noted also that the model is relevant in the assessment of the externalities on property values. Relative to mass appraisal of properties, the model has a long history of use among academics and practitioners (Zurada, Levitan & Guan, 2011: 350). However despite its extensive used, the method is fraught with a number of limitations including inability to handle specification error exacerbated by nonlinearity, multicollinearity and functional form (Do & Grudnitski, 1992: 38; Worzala, Lenk &Silva, 1995: 185). The study of Bourassa *et al*. (2010: 139) having identified these limitations gave a caveat that should follow the use of HRM including careful selection and measurement of relevant variables and ensuring independence of errors one from another.

Recent events in the real estate market have shown a gradual increase in the emergence of several alternative methods used in the assessment of property prices (Lin & Mohan, 2011: 224). This development is noticed in the shift in emphasis towards the additive nonparametric regression (ANR) (Lin & Mohan, 2011); support vector machines (SVMs) (Lam, Yu & Lam, 2009; Zurada *et al*., 2011); artificial neural networks (ANNs) (Borst, 1991; Worzala *et al*., 1995; Zurada *et al*., 2011; Lin & Mohan, 2011; McCluskey, McCord, Davis, Haran, & McIlhatton, 2013); M5P trees (Zurada *et al*., 2011) and spatial and temporal models including geographically weighted regression (GWR); geographically and temporally weighted regression; simultaneous autoregressive model (SAR) (Pace, Barry, Gilley & Sirmans, 2000; Sun, Tu & Yu, 2005; Huang, Wu & Barry, 2010; McCluskey *et al*., 2013 and Fotheringham, Crespo & Yao, 2015) amongst others. While some of these models have their root in the HRMs (global and local), others have completely different underlying philosophies. However, it has generally been established that all models have their individual strengths and weaknesses (Kauko & d'Amato, 2008: 17; McCluskey *et al*., 2013: 240) and there has not been a general consensus on any of the appraisal techniques. Again McCluskey, Davis, Haran, McCord & Mcllhatton (2012: 275) observed that there is no generally accepted class of nonlinear models that can be applied to explore multivariate relationships because of the myriads number of available models.

Prior studies have concentrated on comparing the predictive performance of hedonic regression and ANN models (McCluskey *et al*., 2012; McCluskey *et al*., 2013), hedonic regression and two or more of the following models such as ANNs, GWR, SAR, SVM, M5P trees, ANR, conditional autoregressive and multilevel models (Bourassa *et al*., 2010; Zurada *et al*., 2011; Lin & Mohan, 2011; McCluskey *et al*, 2013; Feng & Jones, 2015) but very little is reported about comparing these models in a single study. Moreover, there is hitherto not a study in the South African context that capture the improvements made to some of these models. For instance, the HRMs have improved versions referred to as global and local regression with global dominating (Fotheringham *et al*. 2015: 418). The improved models are designed to tackle the problems of spatial dependence and spatial heterogeneity in property price modelling. The global models are specifically designed to tackle the problem of spatial dependence while local models generalised the global by allowing model coefficients to vary over space so that spatial heterogeneity are accounted for in the process. Moving away from the improvement made to regression based models to handle spatial effects is the ANN based techniques and other non- or semi-parametric regression designed to deal with the functionality issues of the hedonic regression models. The ANNs is a semi-parametric regression model that provides a useful alternative to traditional hedonic regression because of its ability to efficiently model relationships among nonlinear property data (Peterson & Flanagan, 2009: 148). The most widely used ANN training algorithm is the back propagation. There are other relatively unexplored training algorithms including conjugate gradient (scaled conjugate gradient and Powell-Beale conjugate gradient) and Levenberg-Marquardt in the mass appraisal environment. There are other high powered algorithms that have ability to undertake global search for attribute weights and training of ANNs including the cuckoo search (CS), genetic algorithm (GA) and particle swarm optimisation (PSO).

Therefore the main concern in this study is to investigate into the methods that have been used to model property prices with the purpose of comparing performance and predictive accuracy. Also improved methodologies and algorithms are proposed to enhance the predictive performance of models with obvious limitation(s).

1.2 Problem Statement

Mass appraisal of properties has continued to attract worldwide attention because of the ease and cost effectiveness that follow the assessment. The outcomes of such assessment are widely used in government for property taxation, in business for mortgage underwriting and dissolved business for asset distribution to creditors and owners. Traditionally, the most popularly used techniques in the assessment of property prices is the HRMs. The techniques have however been widely criticised because of its parametric restrictions and a *priori* assumption about the type of data relationship (such as linearity). Because there is no specific form that the relationship should take, various explorations are made to determine the appropriate functional form of model. Moreover a wrong choice of functional form could leads to different conclusion that have catastrophic consequences to the property market players.

Accordingly the study of McCluskey *et al*. (2012: 274–275) report that the choice is between applying linear (or log-linear) models, such as the Box-Jenkins transfer functions and vector autoregressive models. Again, exploring the multivariate relationship that exists within a property data has not been generally related to a particular class of nonlinear models. This is because over the years so many alternative models that have capabilities of handling nonlinear relationships including neural networks, support vector machines, additive nonparametric regression, M5P trees have been developed. There are spatially weighted regression models that have also been explored in handling both property attributes and spatial effects, thus tacking the problem of spatial dependence and spatial heterogeneity. Since neural networks and other non/semi-parametric models are basically nonlinear statistical techniques they offer a good platform for a comprehensive statistical analysis of the problem.

However, Peterson & Flanagan (2009: 150) observe that critics of the neural networks also cite the relative ease of interpretation of HRMs; in particular, partial differentiation of linear models easily isolates each explanatory variable’s is more difficult given variable interdependence, it is relatively straightforward to uncover individual variable attribution (Garson, 1991; and Intrator & Intrator, 2001). Thus, while the contribution of, say, square footage, to property price in a neural network cannot be reduced to a single beta, it can nevertheless be assessed by other means (e.g., simulation methods). At any rate, mass appraisal assessment relied primarily on the framework of HRMs despite the problems associated with nonlinearities, non-normality of inputs, and multicollinearity.

1.3 Aim of the Study

The aim of this thesis is to build improved ANNs optimisation and training algorithms, thus enhancing model's capability and to compare its performance with the estimated results of other mass appraisal methods in terms of their predictive accuracy, transparency, stability of output, defensibility and applicability within the mass appraisal environment.

1.4 Objectives of the Study

The aim of this thesis is accomplished through the following objectives:

i. Establish a baseline regression model for the Cape Town property market;
ii. Assess the performance of flexible regression models in mass appraisal of properties;
iii. Investigate the influence of ANNs training algorithms in mass appraisal of properties;
iv. Build a hybrid model from existing ANNs to create a more effective algorithm in mass appraisal/valuation of properties; and
v. Compare performance of the models in terms of predictive ability, explainability, defensibility and used within the mass appraisal environment.

1.5 Research Methodology

The following methodology was used to achieve the aim and objectives of this study.

1.5.1 Literature Review

The literature review was extensively carried out to unrivalled different mass appraisal models including their strengths as well as their weaknesses for possible improvement in the study.

1.5.2 Data

The data used in the study comprises of sales transaction and different property characteristics. The city of Cape Town jurisdiction was selected because in South Africa they have a relatively well inform database. This was used to derive the baseline model for comparison with other models. Depending on the compatibility of different models to the use of dummy, continuous and categorical format, the data was treated to handle all strengths and weaknesses of models.

1.5.3 Establishing a baseline regression model

To establish a baseline model for Cape Town property market takes some procedures. First, the issue of neglected nonlinearity (see detail in Peterson & Flanagan (2009); and McCluskey *et al*. (2012)) is considered to support the application of nonlinear models. Secondly, the original data was transformed into a set of variables suitable for calibrating a model. In building the baseline model, a consideration was made on the following three alternatives. The first is the linear model; the second is the linear model with the natural log transformation on the dependent variable, referred to as the semi-log model and finally, the third is the natural log transformation of both the dependent and independent variables which is referred to as a log-log or double log model.

1.5.4 Assess the performance of flexible regression models in mass appraisal of properties

There are a number of other techniques designed to deal with the parametric restrictions of the HRMs. These techniques are non- or semi-parametric regression models including geographically weighted regression, simultaneous autoregressive, support vector machines, additive nonparametric and M5P trees. The models were all tested against the industry based techniques and other accuracy test statistics to ascertain their suitability in mass appraisal using the Cape Town property data. The property data was structured in such a way that reflects compatibility with the models. When used for the spatial models, the x, y coordinates were calibrated. The data was stratified into 70% training, 30% testing and the 100% whole set for the other models.

1.5.5 Test the influence of ANNs training algorithms

This research uses four training algorithms including scaled conjugate gradient (SCG), Levenberg-Marquardt (LM), Powell-Beale conjugate gradient (PBCG) and BP. Different network parameters were used to achieve optimal results implemented in Matlab R2013b and WEKA 3.6. Again in modelling property prices in this activity, the data was split into 70% training and 30% testing sets used for modelling and generalisation.

1.5.6 Build a hybrid model from the existing artificial neural networks

The existing BP-ANNs training algorithm is combined with meta-heuristic algorithms namely GA and PSO to search for attributes weights from global space and train the ANNs for improved prediction of property prices. The RapidMiner Studio 7.4 was used for the analysis. The results are also compared with the standalone ANNs in this study.

1.5.7 Compare performance of models in terms of predictive ability, explainability, defensibility and used within the mass appraisal environment.

The purpose is to compare performance of different models with the baseline model in 1.5.3 above. In this study, different mass appraisal models including hedonic regression model, support vector machines, additive nonparametric regression, M5P trees, geographically weighted regression, spatial lag and spatial error models. Various improvements to the existing model structures are effected on deserving models before comparison. The models are ranked in terms of their predictive accuracy, transparency, stability of output, defensibility and applicability within the mass appraisal environment.

1.6 Contributions to Knowledge

This thesis contributes to knowledge as follows:

i. Analyse the significance of spatial variation of property prices, with Cape Town, South Africa used as case study;
ii. Build a hybrid system of PSO and BP in mass appraisal;
iii. Improve the training of ANNs in mass appraisal with SCG and PBCG algorithms; and,
iv. Extend the use of GWR, SEM, SLM, SVMs, ANR and log transformation of variables into the South African property market context.

1.7 Organisation of the thesis

This thesis is organised into six chapters. The structure is as follows:

Chapter one provides an overview that comprises the introduction and background leading to the study, the problem statement, aim and objectives of the research, and the methods used in this thesis.

Chapter two gives the general review of related literatures. First, it reviews previous studies on modelling property prices with the HRMs; secondly a review of some recent development in property prices with spatially varying and weighted regression. In particular, previous studies of modelling prices with geographically weighted regression; simultaneous (spatial lag and error models) is presented. Also previous studies on the use of additive nonparametric regression in modelling property prices were reviewed. This is followed with review of studies on M5P trees and the support vector machines, and lastly, the review focused on modelling property prices with the ANNs.

Chapter three introduces the modelling techniques used in this study. Details about the models and their underlying philosophies, strength and weaknesses are provided. The models are the HRMs; closely followed by the spatially varying and weighted regression models; the ANR and M5P trees; the SVMs, ANNs and hybrid systems. The training algorithms used as first attempt to train ANNs in mass appraisal are provided in this chapter. Furthermore the PSO is introduced and how the proposed hybrid systems will work in weight optimisation and ANNs training in this study.

Chapter four starts with an explanation on the data source and how the variables are chosen for this analysis. This is followed by data cleaning, conversion from text to numeric and removal of extreme and unreliable transactions from the data. The process of accounting for spatial effects and time trends was also carried out in this chapter and finally performance measures used was discussed.

Chapter five commence with the reports of the empirical modelling results to select a baseline regression model for the Cape Town property market. The modelling performance of the spatially varying and weighted regression and all other techniques in prediction exercises is also given in this chapter. Lastly, a comprehensive comparative analysis of models that emerged best in their respective category was performed in order to select the overall best model for the Cape Town property market.

Chapter six summarises the findings based on the realisation of the study objectives, the conclusion, recommendations for future research and practical application of the study is given.

2. CHAPTER TWO

2.0 Literature Review

2.1 Introduction

Mass appraisal is the estimation of market values on a number of properties over a given time period using standardised techniques (d'Amato & Kauko, 2008: 280). These techniques, especially the HRMs, have been integrated as a part of CAMA systems by the property tax assessment community (Eckert, 1990). According to Thompson (2008: 28) CAMA and modelling techniques became popular in the USA during the 1970s and 1980s due to the efforts of various individuals at the Lincoln Institute of Land Policy and the International Association of Assessing Officers (IAAO). Accordingly there exist a number of studies that utilise CAMA models to estimate property prices. The purpose of this chapter is to compile literature concerning property price modelling with different techniques. The chapter will review different techniques; their effectiveness in modelling prices and what this current study intends to contribute particularly as it concerns the geographical context the study has undertaken.

The chapter is divided into three main parts including reviewing studies that utilise the traditional hedonic regression modelling in property price estimation, the second category reviews previous studies on the ANNs and compare their performance with HRMs and other mass appraisal models and lastly review of studies that combined two or more appraisal techniques for pricing of real property. The study of Borst (2007: 24) noted that regression and the ordinary least squares method (OLS) play an important role in the development of literature on property pricing. Therefore a study such as this will commence with a review of literature on studies (traditional regression) that prelude the application of ANNs and other techniques in modelling property prices.

2.2 Pricing of properties with traditional hedonic regression

According to Goodman & Thibodeau (1998: 122) the hedonic modelling has given crucial procedures for analysing commodities that had previously seemed extraordinarily complex. The real property fell under the category of previously hard to analyse commodities but the work of Rosen (1974) provided a framework for real estate pricing. The central idea behind the hedonic modelling or the HRM is that different unit (attributes) of a property are aggregated to develop a price for the property. The use of this technique can be traced to Court (1939) who used it in the automobile industry and also Lancaster (1966) who relates it to the bundle of characteristics which provide utility to consumers of goods. The study of Malpezzi, Ozanne & Thibodeau (1980) relates real property to consumer goods like a bundle of groceries having different sizes and items. Relating this to the different component features of property, Sirmans, Macpherson & Zietz (2005: 4) reported that property is a bundle of structural, environmental and spatial attributes. These attributes differ one from the other property, therefore the HRMs would price the features on a collective sample of many dwellings.

The study of Des Rosier & Thériault (2008: 113) listed a number of areas that the model had previously been used and found the model to particularly be important to the property market because of high level of competition that is open to buyers and sellers. In practice the model has been used for pricing of real property for more than four decades. Accordingly, Goodman (1978) used the HRMs on a data base of 1835 single family dwellings in the New Haven, standard metropolitan statistical area (SMSA) to form submarkets and indices that determine relative prices of housing services. However, despite the robustness of this model the choice of functional form and ability to capture spatial features are amongst its limitations. On functional form specification, the economic theory fails to specify the functional relationship that a property price and its attributes should take.

This according to Crooper, Deck & McConnel (1988: 668) led a number of scholars to rely on goodness of fit criterion, as suggested in Rosen (1974) and Goodman (1978), to select appropriate functional form. However, their study noted that when all attributes are observed, linear and quadratic Box-Cox transformed variables gives accurate estimates of marginal attribute prices but this change when certain variables are not observed or replaced by proxies a linear function outperformed the quadratic Box-Cox function. Again, the study of Goodman (1978: 483) found that the frequently used linear form to be overly restrictive and favour the Box-Cox transformation. Several criticisms trailed the use of Box-Cox function leading to a situation where some authors directly formulating a model structure without recourse to the hedonic function (Borst, 2007: 35).

In general, apart from the basic linear multivariate specification, authors commonly used the semi-log and log-log specifications. In the semi-log formulations, the left hand side of the equation namely the dependent variable (property price) is regressed against the linear arrays of structural and locational characteristics while in the log-log formulation, both the dependent variable (property price) and independent variables (structural features) are transformed. Therefore, all assessment must at least take a particular form which researchers in practice usually specify in their model calibration. Accordingly Kang & Reichert (1987: 1) noted the significance of regression coefficients and that the prediction accuracy depends on the choice of estimating technique and the functional form of the regression equation. The study of Schulz, Wersing & Werwatz (2014) used the price and log-prices as dependent variable while other variables remain in their linear format on 18,444 single-family transactions to estimate property prices and found the semi-log model to be a better choice. Also the study of McCluskey *et al*. (2012) undertook a study on a 2694 property dataset in the Lisburn district of Northern Ireland, UK using linear, semi-log and log-log HRMs and found the semi-log model to outperform other models. The three studies show that log transformation could improve the predictability of HRMs. However a dissimilar result was achieved with semi-log model, in a study by McCluskey (2016) on a dataset of 46,689 (before the removal of outliers) in Kazakhstan using three scenarios that included analysis after the first data cleaning, second data cleaning and third semi-logarithm. The best result (R^2 = 67%) was obtained after the second data cleaning with linear additive regression (linear model).

The result of reviewed studies indicates that log transformation is not always assured of optimal results. Therefore it is good to first undertake a test that supports the need for log transformation of variables before embarking on mass appraisal assessment with the HRMs. There are other studies that improve on the parametric nature of HRMs, particularly as it concerns the functional form specification, spatial heterogeneity and spatial dependence. The next section would review these studies.

2.2.1 The use of flexible regression techniques in modelling property prices

The word “flexible” as used here is taken to mean HRMs that are devoid of the many parametric restrictions. There is a growing interest in property pricing with a number of flexible approaches including artificial neural networks, fuzzy logic, support vector machines, rough set theory, additive nonparametric regression, geographically weighted regression, simultaneous autoregressive amongst others exist. According to Kauko (2003: 254) the term “flexible regression” was used by Verkooijen (1996) rather than non- or semi-parametric regression. Kauko (2003: 254) further made a distinction between the parametric regression models and non- or semi-parametric to include low variance and high bias, whereas, the non- or semi-parametric generate greater variance and lower bias. In this section, the concern is to review previous works on the regression based flexible models. Several studies have been undertaken with non-parametric and semi-parametric regression to estimate hedonic prices. The study of Anglin & Gencay (1996: 634) reported that one of the ways of making a regression function robust is to use a model that does not have an *a priori* parametric restriction such as a nonparametric regression. The study used 546 transactions from a dataset supplied by the Windsor and Essex County and found that the ANR model permits random interaction between many more of the regressors than the parametric Box-Cox model. Again in a related study, Gencay & Anglin (1996) used the same data source having 955 property transactions sold in Windsor in 1990, the result shows the ANR model (mean absolute error $1230) to outperform the three parametric models (average mean absolute error $2790).

The non-parametric or semi-parametric have proved useful in smoothing the parametric restrictions of the HRMs, thus enhancing the predictive accuracy as seen in the reviewed studies. In addition to the studies above, Bin (2004) used a total of 2,596 single-family property sales from Pitt County, North Carolina and found the ANR model to outperform the parametric models in both in-sample and out-sample price predictions. Recently, Lin & Mohan (2011) did a study in Amherst, New York using 33,342 single-family transactions and found the ANR model to perform well in prediction. The nonparametric regression incorporates the nonlinearity among variables that are difficult to be captured in the parametric regression. Besides the use of ANR that avoid the parametric restrictions of the HRM, Goodman & Thibodeau (1998) used hierarchical linear modelling that shows price as a function of consumer choice and interaction between property and neighbourhood characteristics and submarkets. There are other semi-parametric regression models that are designed to capture spatial effects and model property prices with great precision.

Des Rosier *et al*. (2008: 134) observed that in hedonic regression the contextual variation over space are usually specified with “fixed” coefficients derived from locational dummy variables to assess their direct effect on property prices. However, this stable price assumption is not possible where the markets are heterogeneous making it imperative to account for spatial effects within the regression framework (Theriault, Des Rosier, Villeneuve & kestens, 2003). To tackle spatial heterogeneity, the spatial expansion method pioneered by Casseti (1972; 1997) and GWR are among others used. Accordingly the study of Kestens, Thériault & Des Rosier (2006) using transaction data of 761 single-family dwellings in Quebec City, Canada, employed the spatial expansion method and GWR and found that both methods provide conclusive results relative to the marginal value given to property and locational attributes based on the characteristics of the buyers' household. Farber & Yeates (2006) used the OLS and compare performance with GWR, spatial lag model and moving window regression on a data of 19,007 freehold housing sales in Toronto, Canada and found the GWR to be less spatially biased, account best for the spatial variation in prices but did not support its adoption by assessment community because of limitations in the statistical framework used. Bitter, Mulligan & Dall'erba (2007) used spatial expansion method and GWR to account for the spatial heterogeneity and prediction of property price in Tucson. The study used the transaction data of 11,732 single-family residences and found that both methods improve the results of stationary coefficient models but in terms of explanatory power and predictive accuracy the GWR performs better. Also McCluskey & Borst (2011) used the GWR with datasets from Catawba, Sarasota and Fairfax Counties in the USA to identify market segments and conclude that the resultant segments have propensity of improving predictive accuracy and lowering spatial autocorrelation in the residual errors. Again, the study of McCluskey *et al*. (2013) used the GWR to assess property prices in Northern Ireland, UK. With a sample of 2,694 residential properties, the GWR was found to outperform the other models used for comparison in terms of cost efficiency, ease of use, and predictive accuracy.

According to Valente, Wu, Gelfand & Sirmans (2005: 109) spatial effects (dependence) are being modelled by spatial conditional autoregressive (CAR), spatial simultaneous autoregressive (SAR) and kriging (Pace, Barry & Sirmans, 1998: 8–9) modelling. Spatial dependence or autocorrelation occurs when there is interdependence among observations in a geographical space that violates the assumption of uncorrelated error terms (Osland, 2010: 290). The SAR has been used to account for spatial autocorrelation in a number of studies. For example, the study of Pace & Gilley (1997) used 506 housing datapoints from Boston SMSA and found SAR to outperform the HRMs. Dubin, Pace & Thibodeau (1999) used a small sample (10) to compare the regression coefficients of OLS and four spatial techniques including SAR, CAR, mixed regressive spatially autoregressive and Gaussian correlogram. The study used the property selling price as a function of square feet of living area and location and found that apart from the mixed regressive spatially autoregressive model, other spatial models yield parameter estimates that are slightly closer to the true values than the OLS. Wilhelmsson (2002) used 1,377 single-family property transactions in the municipality of Stockholm, Sweden and found the autoregressive model to outperform the traditional hedonic regression. Militino, Ugarte & Garcia-Reinaldos (2004) used the OLS, lattice, geostatistical models on a 293 property dataset in Pamplona, Spain and found that the lattice models (SAR and CAR) coped with the spatial dependence and gives a robust inference similar to other spatial techniques used. Bourassa, Cantoni & Hoesli (2007) used a sample of 4880 residential property sales in Auckland, New Zealand and found the lattice models (SAR and CAR) to offer less prediction accuracy than the OLS and geostatistical (exponential and robust exponential variogram) models. Again, McCluskey *et al*. (2013) used SAR in the Northern Ireland, UK and found it to perform well. The benefits of using SAR and CAR are that it gives a nearest neighbour based smoothing of the means, convenient computation and improve the explanation of error

(Valente *et al*., 2005: 110). The limitations of these techniques are in their inability to directly model location and build in spatial prediction.

The reviewed studies show that although HRM is widely acceptable for use in property related businesses, it has many weaknesses which has since been identified and addressed through various improvements. This current study will extend the use of these techniques into the South African property market context because such improvement namely using spatial modelling techniques and ANR do not exist. In modelling hedonic price function with the ANR, the central idea is to replace the commonly used linear function with an unspecified smooth function while keeping additive framework of the linear regression models. Although significant improvement has been made to the HRMs, there are other studies that suggested the need to apply the artificial intelligence techniques into the mass appraisal environment (Do & Grudnitski, 1992; Tay & Ho, 1992).

2.2.2 Modelling property prices using support vector machines and M5P trees

Other techniques that have recently been introduced into the mass appraisal environment include SVMs and M5P trees, with little research undertaken on these. The SVMs was first mentioned as a technique for mass appraisal in Lam, Yu & Lam (2009). The study used 4,143 and 21 property transactions in two case studies of Hong Kong and Nanjing, Mainland China, respectively and found the SVMs to outperform ANNs and HRMs in factor weighting and predictive accuracy. Again, though mentioned previously in this study, Zurada *et al.* (2011) used the SVMs for mass appraisal in the USA and found it to be among the best performing models. The SVMs used different kernel function to construct a separating hyperplane in high dimension feature space without plainly performing the computation in the feature space (Zurada *et al.*, 2011: 362), but the choice of kernel function is not a straight forward procedure. Cui & Curry (2005: 608) reported that there is no complete working meta-theory to assist with the selection of kernel transformation for SVMs. Therefore a high reliance on trial and error process is prevalent until a kernel that provides optimal result is guaranteed. The implication of the trial and error procedure is that a kernel that works well with a particular class of data or market context might not work well with another. The choice of a kernel function is seen in the studies of Lam *et al*. (2009) and Zurada *et al*. (2011). Whereas the study of Zurada *et al.* (2011) used a polynomial kernel function for their analysis, Lam *et al.* (2009) utilised radial basis kernel function (RBF) in their analysis. In this study the polynomial, normalised and redial basis kernel functions are used.

The M5P trees like SVMs, is not popular in the mass appraisal industry. Like other nonparametric models, the technique does not require assumption about the dependent variable distribution (Acciani, Fucilli & Sardaro, 2011: 28). The only known studies that utilised the M5P trees for mass appraisal are Zurada *et al*. (2011) and Acciani *et al*. (2011). In the study of Acciani *et al*. (2011) 4 market segments (submarkets) were created on a dataset of 169 trading instances of trullo-inclusive farms in the Ceglie Messapica, Cisternino, Fasano and found M5P trees to possess a higher statistical significance in the prediction of prices than the HRMs but marginally below the multivariate adaptive regression splines. The model was also found to perform well in the study of Zurada *et al*. (2011). Having observed the promise these models have, coupled with the relatively sparse literatures in the mass appraisal environment, the current study will investigate their use in the South African property market as a first attempt.

2.3 The application of ANNs in mass appraisal of properties

The ANNs have been applied for mass appraisal since the 1990s in many research papers, but with mixed results (McCluskey, *et al*. 2012: 275). Several papers have been written on the potential of the model within the real estate sector. While some studies found the model as a useful tool for effective prediction of property values others took a contrary position. For instance, earlier studies of Borst (1991), Evans, James & Collins (1992), Tay & Ho (1992), Do & Grudnitski (1992) found it useful in mass appraisal. To support this position, a comparison was made in some studies between ANNs and HRMs in different countries. Tay & Ho (1992) in Singapore; Do & Grudnitski (1992) in the USA; Evans *et al*. (1992) in the UK; Borst (1995) in the USA, all found ANNs to outperform the HRMs. Similar findings were made in other studies such as Nguyen & Cripps (2001) in the USA; Limsombunchai, Gan & Lee (2004) in New Zealand; Peterson & Flanagan (2009) in the USA. However, the studies of Worzala, Lenk & Silva (1995), Lenk, Worzala & Silva (1997), McGreal, Adair, McBurney & Patterson (1998) cautioned the appraisal community against its use because of inconsistent results found in their studies.

Recent studies incorporated other models into the mass appraisal industry. The bottom line is to test every model against IAAO benchmarks and other statistical accuracy tests to determine their suitability. Accordingly Zurada *et al*. (2011) compare HRMs, ANNs, additive regression (AR), M5P trees, support vector machines with sequential minimal optimisation (SVM-SMO), radial basis function neural networks (RBNN) and memory-based reasoning (MBR) in Louisville, Kentucky, USA. The result reveals that non-traditional regression based models (AR, M5P trees, SVM-SMO) performed better in all of the five simulated experiments, particularly with homogenous data, while artificial intelligence (AI) based models (ANNs, RBNN and MBR) performed better with less homogenous datasets. Lin & Mohan (2011) in Amherst, New York, compare the predictive accuracy of HRMs, ANNs and ANR and found ANNs to consistently perform better than HRMs and ANR in both training and testing/validation sets. McCluskey *et al*. (2012) in Lisburn, Northern Ireland investigated the predictive abilities of ANNs and three regression functions (linear, semi-log and log-log) and found the three regression models to perform better than ANNs. Furthermore, McCluskey *et al*. (2013) compare the predictive accuracies of HRMs, SAR, GWR and ANNs and found GWR to outperform all other models.

The studies of Peterson & Flanagan (2009) and McCluskey *et al*. (2012) are of significant interest to this current study because of the RESET test carried out to support the case for the use of the nonlinear models. Peterson & Flanagan (2009) observe that the HRMs are exposed to pricing errors owing to the way means are extrapolated from large samples and as such will also be exposed to significant sampling errors. The study also noted that specification error is also unavoidable in *ad hoc* specification and to the extent that value does not map linearly onto property characteristics, so too are errors due to neglected nonlinearities. Again to the point that nonlinear models integrate linear forms, then nonlinear models would be the desired choice. However, the exact nonlinear form is neither apparent nor are there practical steps one would take to find the correct form. The ANNs do, however, provide a practical alternative to conventional least squares form (including nonlinear least squares) that is easily implementable and which efficiently models nonlinearity in the underlying relationships (including the parameters) according to Peterson & Flanagan (2009).

The foregoing studies provided conflicting results which show that some models performed better in certain accuracy and benchmark tests but poorly in others. The study of McCluskey & Anand (1999: 221) provided a summary of different techniques highlighting their strengths as well as weaknesses and concluded that “most, if not all” models have weaknesses in terms of predictive accuracy, transparency, and ease of application. Kauko & d’Amato (2008: 17) noted this and suggests the need to reflect on pertinent issues. For instance, is it on the accuracy of models, the feasibility or some non-technical part that might involve more detailed analysis such as the adjustment of the structure, or that when these issues are reflected upon, some improvements or modifications should be implemented so that optimality could be realised in a model(s). Such improvement may be effected on the model structure or data quality.

Accordingly, McCluskey *et al*. (2012) utilised the different regression models having observed the limitations of HRMs in handling nonlinearity in property data relative to the ANNs. The current study extends their work by improving on the structure of the ANNs in the field of mass appraisal. The ANNs have a structure by which the weights of the links between the processing elements are adjusted on the basis of patterns in the property dataset. Should discrepancies occur, the weights are altered changing the original state of the network, so the system appears to learn (McCluskey, 2012: 275). This alteration is performed with the aid of the back propagation (BP) learning algorithm. The BP was first developed by Werbos (1974) and popularised for multilayer perceptron by Rumelhart, Hinton & Williams (1986). BP uses the gradient descent search method to modify connection weights in order to minimise error between actual and desired output vectors. The algorithm is simple in execution and suitable for solving different problems. However, foremost among its challenges are inability to continuously achieve global optimum and possibility of getting entrapped (not able to complete the training cycle) into local optima. To enhance ANNs performance, meta-heuristic algorithms such as genetic algorithm (GA), cuckoo search (CS), particle swarm optimisation (PSO) etc. were introduced to optimise and train ANNs, in order to speed up the rate of convergence and escape from being entrapped into local optima. These meta-heuristic algorithms have been used to optimise and train ANNs in many fields but PSO is yet to be used for mass appraisal of properties.

2.3.1 Combining models for mass appraisal of properties

Combining algorithms in real estate assessment has been reported in literature. McCluskey & Anand (1999: 221) and Kilpatrick (2011: 538) suggests that combining two or more techniques should achieve enhanced capabilities of the new hybrid model. For example, the study of Bourassa, Hamelink, Hoesli & MacGregor (1999: 161–162) used principal component and cluster analysis to assess housing submarkets in Sydney and Melbourne, Australia. They first used principal component analysis (PCA) to extract important factors from the dataset and secondly used cluster analysis (CA) to form the most appropriate submarkets. Again after achieving these, they performed multiple regression assessment for the submarkets that were formed to estimate hedonic price equation in each city. Gonzalez & Formoso (2006) built fuzzy rule based on gross building area (GBA) and fuzzy rule based on location in Porto Alegre, Brazil. In building the fuzzy rule of the former, each membership function is defined using the limits of GBA while the genetic algorithm was used to estimate the fuzzy rules. The fuzzy rule based on location was constructed in similar fashion but have differences in the procedures utilised. The rules are determined in a specialised manner according to the region of the city but all contribute to the final estimates. Similarly, Guan, Zurada & Levitan (2008) combined fuzzy systems with neural networks to assess property prices in the Midwest region of the USA. While the fuzzy system was used in generating fuzzy rules and parameters for the membership function, the neural network was used to fine tune the fuzzy rules and found results that are comparable to the regression model.

Also McCluskey & Anand (1999) utilised 412 property transactions to build a hybrid structure with GA and back propagation ANNs in collaboration with *K*- nearest neighbour (*k*-NN) algorithm. The nearest neighbour algorithm was used to enhance the selection of comparable properties and give the appraisal process transparency and explainability. In determining attribute weights for the network, three techniques which include: deriving weights from the domain expert; discovering attribute significance and transforming them into weights via a loosely coupled neural network and using a tightly coupled genetic algorithm for discovering attribute weights. The result shows that the tightly coupled genetic algorithm achieved the lowest mean absolute error (MAE) of £4,785; this contrast the high MAE of £9,842 achieved for the baseline model. The result also reveals the back propagation trained ANNs with three layered architecture achieved a MAE of £4,884 (see table IX of their study). The study noted that in terms of predictive accuracy, the technique performed optimally but became affected by the "black box" non-transparent nature of the entire process. A comparison of the result reveals that the hybrid neural network with Euclidian distance metric performed abysmally below the standalone ANNs even though the process was transparent. The simple explanation adduced in McCluskey & Anand (1999: 232) was that genetic algorithm uses the biases of the nearest neighbour algorithm to arrive at attribute weights, while the ANNs do not. The two algorithms (GA and *k*-NN) have different ways of considering the significance of attributes. The current study will use the PSO and GA independent of *k*-NN to select attribute weights relative to the domain expert selection of weights for the standalone BP-ANNs.

However, the "black box" non-transparent nature of the ANNs is a notable concern within the mass appraisal environment. ANNs are no longer a new concept within the assessment appraisal community, but the IAAO (2003) observe that the problem with the concept is the non-apparent working mechanism and difficulty of explaining the model structure. In a recent study, Grover (2016: 199) reported that the results of ANNs are something that users have to take on trust. If users are mortgage banks that require the result as a check on requirement for single valuation or updating prior assessment on existing loans, then problems will not ensued. There is a problem if the purpose is for property tax assessment where right to challenge the estimate and make appeal at tribunal is permitted. The transparency of the process leading to price estimate is what puts the HRMs at a distinct advantage despite the misgiving that followed the use amongst others. Therefore, a little measure of transparency is added to the ANNs through weight optimisation with PSO and GA. The search for attribute weight from global space will show the different contribution and relative importance of property attributes to the prices.

2.4 Criteria for the selection of appropriate pricing model

The foregoing studies utilise varying criteria to assess the performance of models. There are limited studies that suggest the use of uniform criteria for the assessment of a model. Accordingly, Thibodeau (2003: 2) suggests the use of predictive accuracy estimate as a measure of testing model's performance. Kryvobokov (2004: 221) advocates five criteria for assessing model suitability in Ukraine including (i) clearness of method; (ii) measurability of the result; (iii) relevance of the result; (iv) market orientation of the method; and (v) simplicity rather than accuracy of the method. These were expanded in d' Amato & Kauko (2008: 293-294) to include the following groups and subgroups of criteria: a. *institutional criteria*, (i) suitability of methodology to the property market context; (ii) specific path-dependence; b. *methodological criteria*, (i) accuracy of independent valuations; external or out-sample validation; (ii) conceptual soundness; (iii) analysis of valuation variation; (iv) internal consistency of the model structure/predictions; internal or in-sample validation; (v) nature of the adjustment; (vi) reliability and robustness of the model; and (vii) feasibility. However, whilst it is imperative to assess the suitability of method(s) using the criteria above, the study of McCluskey (1997); McCluskey *et al.* (2012); McCluskey *et al.* (2013) added the explicit explainability of the method so that appraisers can easily defend the estimates before a tribunal or in a formal court. The ability of a model to provide detail explanation of the appraisal process is germane hence the latter criterion is seen as fundamental to the assessment of model performance/suitability in this study.

2.5 Chapter analysis, summary and conclusion

This chapter reviewed previous research in line with the objectives of this study. The review shows the studies that utilise various methods and what they have achieved. The HRMs received consideration in virtually all the reviewed studies. Depending on what some of the studies have set out to achieve, obvious limitations and strengths of models were noted with some of them identifying and utilising the improvements made to the HRMs. While there are a considerable large number of studies that utilise the HRMs, there exists a scant body of knowledge on the spatial models, additive nonparametric regression, support vector machines and the M5P trees. Very little has been done to predict property prices with these models, particularly, the support vector machines, M5P trees and the hybrid systems.

The artificial neural network modelling approach also receives a number of mentions in the reviewed studies but there has been no mention of building a hybrid system of PSOBP. Though, studies such as McCluskey *et al*. (2012; 2013) partly captured the goal of this study, this research is unique in that it will bring together a number of models into one single study. Again the geographical contexts of these studies reveal a vacuum that requires filling in the South African context. As Bourassa *et al.* (2010: 140) observe, the results of previous studies cannot be implemented in other climes or regions because of differences in the data. Also the findings of Bourassa, *et al*. (2007), that the lattice models (SAR and CAR) offered less prediction accuracy than the OLS despite their enhance performance in other studies, justify the need to test the spatial lag and spatial error models in this study. Therefore this study is undertaken to provide a modelling framework for property pricing in the Cape Town, property market.

3. CHAPTER THREE

3.0 Mass Appraisal Modelling Techniques

3.1 Introduction

According to Acciani *et al*. (2011) data mining is a recently developed field (Han & Kamber, 2006) that combines statistical analysis, computer science, artificial intelligence and database management. This field concern with the selection, exploration and mining process of knowledge from multitudes of data, through the application of suitable techniques, that discover possible regularities, trends and associations that are *a priori* not known. The purpose of this chapter is to present and discuss the underlying philosophies of different mass appraisal modelling (data mining) techniques used in this study highlighting their strengths, weaknesses and various improvements (if any) made to surmount all limitations.

The models are HRMs, ANNs, SVMs, and M5P trees. Others that were introduced to improve the existing model structure of hedonic regression are GWR, ANR, simultaneous and conditional autoregressive models. Also algorithms used to improve the ANNs optimisation and training such as cuckoo search, genetic algorithm and particle swarm optimisation are also presented. Lastly, other training algorithms that have been suggested in the literature to have high computing powers including scaled conjugate gradient, Powell-Beale conjugate gradient and Levenberg-Marquardt are presented in the chapter.

3.2 Hedonic Regression Models

According to McCluskey *et al.* (2012: 277); McCluskey *et al.* (2013: 246) the hedonic modelling or regression model has been accepted as the most widely applied technique within the *ad valorem* assessment process. Janssen & Söderberg (1999: 361) report that hedonic modelling gives the framework for the assessment of "differentiated goods like housing units whose individual features do not have observable market prices". The wide acceptability of this model within the property tax environment stems from its ability to explicitly define the entire appraisal process and predictive accuracy. In the context of mass appraisal, hedonic regression models the relationship between property price and its characteristics so that through the interaction, price can be estimated. Additionally, McCluskey (2016: 127) reported that the model measures the different contribution of property characteristics to price and give a weighting for each characteristic. However as noted in McCluskey *et al*. (2012: 278) three important elements must be addressed for effectiveness of this model, which is careful selection of dependent and independent variables; choice of functional form and lastly the statistical relevance and contribution of the independent variables to the model.

In this study, the dependent variable used is the assessed value of properties while the independent variables were selected based on the *a priori* expert knowledge that they are likely to influence property prices and theory. Since economic theory does not specify a particular functional form to use in hedonic regression modelling, the linear, semi-log and log-log models were utilised in this study. The study of Palmquist (1979: 442) pointed out that selection must be approached from an empirical point of view. Moreover advance in technology has been able to address the model complexity through the use of semi and nonparametric techniques. The linear model is given in equation 3.1:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_s x_s + \varepsilon$$

or

$$Y = b_0 + \sum_{i=1}^{s} b_i x_i + \varepsilon \tag{3.1}$$

where Y is the assessed value (dependent variable); x_1, x_2, …, x_s are the independent variables; b_1, b_2, …, b_s are coefficients or prices per unit assigned by the algorithm to the independent variables; b_0 is the regression constant and ε is the error term. To handle a nonlinear interaction between variables the semi-log and log-log models given in equations (3.2) and (3.3) are used:

$$LogY = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_s x_s + \varepsilon \tag{3.2}$$

and

$$LogY = b_0 + b_1 Logx_1 + b_2 Logx_2 + \cdots + b_s Logx_s + \varepsilon \tag{3.3}$$

The study of Schulz, Wersing & Werwatz (2014: 133) observe that using the log transformation could partially rectify the inherent heteroscedasticity of property prices. It also makes the effective relationship among variables nonlinear while still preserving the linear model. Of the three models (linear, semi-log and log-log) the linear additive are commonly used for single-family dwellings, easiest to calibrate and it also allows the contribution of each variable to be added (Gloudemans, 2002: 26). However, the presence of spatial dependence or autocorrelation and spatial heterogeneity or non-stationarity inherent in the property markets are fundamental issues that inhibit the effectiveness of the HRMs. A simple way of dealing with the spatial and temporal effects is to introduce indicator variables into the model but difficulty arises when the number of indicators is large (Pace, Barry, Clapp & Rodriquez, 1998: 15–16). The challenges are tackled in the improved versions of the HRMs. Fotheringham *et al*. (2015: 418) reported that two versions, the global and local HRMs are used to account for spatial heterogeneity and spatial dependence in property price prediction. Examples of the local models are the moving window regression, GWR and multilevel modelling (Fotheringham *et al.*, 2015: 419) while examples of the global models are spatial autoregressive (SAR) and spatial moving average (SMA) (Anselin, 2003: 313). The techniques of interest in this study are GWR and spatial autoregressive models. These have been applied in some jurisdictions to identify potential spatial variations in the property dependent and independent variables, measure the relationship and improve accuracy of predictions (Tu, Sun & Yu, 2007; Bourassa, Cantoni & Hoesli, 2007; Huang, Wu & Barry, 2010; McCluskey *et al.*, 2013; Fotheringham *et al.*, 2015) but such studies are rare in the South African context.

3.2.1 Geographically Weighted Regression

According to Brunsdon, Fotheringham & Charlton (1996: 284) the GWR is a technique that has its root embedded in the framework of HRMs, which permit its estimates to vary on specific location. The standard formulation is as follows:

$$y_i = b_0(u_i, v_i) + \sum_c b_c(u_i, v_i) x_{ic} + \varepsilon_i \tag{3.4}$$

where, (u_i, v_i) are the coordinates of the ith point in space; $b_c(u_i, v_i)$ depicts a set of values of parameters at the ith point and $b_0(u_i, v_i)$ denotes the intercept value. This technique is more flexible than the HRMs because of its non-reliance on the numerous underlying assumptions and rigidity of having a single model that capture relationships in the entire property market. GWR creates a separate regression model at every observation point which permits the estimation of coefficients at every location (Brunsdon *et al.*, 1996: 284–285; Bitter, Mulligan & Dall'erba (2007: 8). The method uses data points to identify properties that are sold within the area of the subject property and measure the distance between them.

Thus the farther the property is from the subject (regression point) the lower the weights assigned. The study of Borst (2012: 536) illustrates how GWR operates in assigning weights that vary with distance from the regression point. The study utilised the peak of the surface as the regression point, thus showing that any point below the surface is assigned weights relative to its position (height) of the surface at that point. Accordingly Huang *et al*. (2010: 385) and McCluskey *et al*. (2013: 249) report that an estimate is achieve from observation relative to distance between points. The parameter estimate $b_c(u_i, v_i)$ for this is given as:

$$b(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y \tag{3.5}$$

where the spatial weighting matrix is given as $W(u_i, v_i)$. The Gaussian function as used by Bitter *et al*. (2007: 15) and McCluskey *et al*. (2013: 249) is used to specify the Euclidian distance *d* between the regression and the observation points and *h* denoting the bandwidth given as:

$$W_i(u_i, v_i) = \exp(-d/h)^2 \tag{3.6}$$

The estimated results of GWR are sensitive to type of bandwidth used. Therefore care must be taken in the selection of bandwidth. Huang *et al*. (2010: 386) noted that two weighting regimes are used; these are fixed kernel and adaptive kernel. The fixed kernel has a varied number of nearest neighbours but the distance is constant. In the adaptive spatial kernel it is the reverse phenomenon where distance varies while the number of nearest neighbour remains constant. The adaptive spatial kernel that allows the bandwidth to vary based on the density of property sales around each regression point is used in this study. Bitter *et al*. (2007: 15) report that adaptive kernel capture different segments of the property market including smaller area with rich data and larger area with sparse data.

3.2.2 Spatial Autoregressive Models

Spatial modelling techniques take the spatial residual information generated by the hedonic regression to improve accuracy of predictions. The study of Dubin, Pace & Thibodeau (1999: 80) observes the role of the spatial techniques in improving the observation's predicted values when there is under-prediction of the properties around the observation in the HRMs. The equation for the autoregressive model is given as:

$$Y = \rho WY + \varepsilon \tag{3.7}$$

where the parameter ρ is the coefficient of autocorrelation, ε is the vector of the error term, W is the weights matrix which according to Borst (2007: 77) has elements W_{ij} that reduces as the distance between properties i and j increases (Besner, 2002: 195). The relationship is expressed in an n x n matrix of the spatial weight (W) which starts by identifying the neighbourhood structure. This usually lead to the formation of a binary neighbourhood matrix N in which $n_{ij} = 1$ when property j is a neighbour to property i. Furthermore, the elements of N are weighted such that closer properties are assigned higher weights and more distant properties are assigned lower weights.

The study of Dormann *et al*. (2007: 613) report that both the simultaneous and conditional autoregressive (SAR and CAR) models incorporate spatial autocorrelation using neighbourhood matrices which specify the relationship between the response values (in relation to CAR) or residuals (in relation to SAR) at each location, i, and those of the neighbouring locations, j (also Cressie, 1993; Lichstein, Simons, Shriner, & Franzreb, 2002; Haining, 2003).

3.2.2.1 Conditional and simultaneous autoregressive models

According to Dormann *et al*. (2007: 614) SAR can take different forms depending on where the spatial autoregressive process is believed to occur (Cliff & Ord, 1981; Anselin, 1988; Haining, 2003). Accordingly, the first SAR assumes that the autoregressive process is believed to occur only in the response variable (i.e. spatially lagged response model) and thus take the term (ρW) for the autocorrelation in the response variable Y, but also the standard term for the independent variables and errors $(X\beta + \varepsilon)$ as used in the OLS models. In the spatially lagged model (SLM), the response variable is related to itself in a particular way (Borst, 2006: 3). The SLM allow for the observed sale prices of nearby properties to influence the dependent variable in the model. Krause & Bitter (2012: 520) reports that the spatially lagged model attempt to capture the spatial dependence in the property market or account for the influence of sales of properties in nearby locations on current property prices. The basic spatially lagged response model (SAR_{lag}) is given in equation 3.8:

$$Y = \rho WY + X\beta + \varepsilon \tag{3.8}$$

The second SAR assumed that spatial autocorrelation can affect both predictor (independent) variables and response variable leading to the addition of another term $(WX\gamma)$ into the model that describes the autoregression coefficients (γ) of the spatially lagged independent (explanatory) variables (WX). The spatially mixed (SAR_{mix}) version is given in equation 3.9

$$Y = \rho WY + X\beta + WX\gamma + \varepsilon \tag{3.9}$$

Another approach to SAR modeling is the spatial error model (SAR_{err}) which can be applied only if there is significant spatial autocorrelation. In the spatial error model the autoregressive process is believed to occur in the error term and neither in the response variable nor in the explanatory (independent) variables. The study of Kissling & Carl (2008: 61) noted that this is usually the case if autocorrelation is not totally explained by the explanatory variables or if the autocorrelation is an integral property of the response variable itself. For the SAR_{err}, the normal OLS model is added by a term (λWu) which denotes the spatial structure (λW) in the spatially response error term (u). The SAR_{err} takes the form

$$Y = X\beta + \lambda Wu + \varepsilon \tag{3.10}$$

In selecting an approach, de Smith (2011: cap. 16, pp. 61) suggested a Lagrangian Multiplier diagnostics test be used. The major weaknesses of SAR and its variants CAR as Wall (2004) observed are in the area of interpretation of the weighting schemes structure.

Of the three SAR approaches, it is the spatial error model (SEM) that is most similar to the CAR because the error lacks directionality. Consequently their outputs/results are similar, though having differing conceptual designs (Rangel, Field & Diniz-Filho, 2011: 48; de Smith (2011: cap. 16, pp. 64). Because of the similarities in results of SEM to the CAR, the SEM was used in this study. The SLM was also used to tackle the spatial dependence between the dependent variable (lag process). The relevance of using the SEM and SLM in the HRMs is to ameliorate the occurrence of biases and inefficiencies in the coefficient estimates (Mueller & Loomis, 2008: 213). Although CAR was not used in this study, the general formulation is nonetheless given in equation 3.11 (Keitt, Bjørnstad, Dixon & Citron-Pousty, 2002: 618).

$$y = X\beta + \rho W(y - X\beta) + \mu \tag{3.11}$$

with $\mu = M(0, S_c)$.

If $\sigma_i^2 = \sigma^2$ for the entire locations i, the covariance matrix is $S_C = \sigma^2(I - \rho W)^{-1}$, in which W is a symmetric neighbour. According to Dormann *et al*. (2007: 614) the CAR can be unsuitable when directional processes are coded as non-Euclidean distances, leading to an asymmetric covariance matrix. When this occur the best option is to switch to the SAR because their W must not be symmetric. Additionally, Keitt, *et al*. (2002: 618) noted that the CAR model simply considers the first-order neighbourhood effects, while the SAR model allows for recursive, higher order neighborhood effects.

3.3 Additive Nonparametric Regression

According to Lin (2010: 68) the ANR was initially suggested by Friedman & Stuetzle (1981) and popularised by Hastie & Tibshirani (1990). The model was developed to remediate the aspect of functional form and poor handling of nonlinear data resulting in significant error during prediction of market values in the HRMs. Pace (1998: 77) opine that wrong choice of functional form specification leads to all sort of disastrous consequences for the traditional hedonic estimator. Lin & Mohan (2011: 226) report that the main goal of ANR is to modify the linear function $b_i x_i$ of the independent variables in equation 3.1 by an unstated nonlinear smooth function specified as:

$$Y = b_0 + \sum_{i=1}^{s} f_i(x_i) + \varepsilon \tag{3.12}$$

where Y is the sale price of property, x_i is a set of independent variables and f_i denotes arbitrary nonlinear smooth functions whose shapes are unrestricted. In the equation every independent variable must make a flexible contribution towards the determination of market values of properties. Because of its flexibility a regression curve must not be pre-determined prior to assessment because ANR provide a regression curve during its operation (Lin & Mohan, 2011: 226). Also, the nonlinear effects of continuous covariate and time trend are modelled through penalised splines, while discrete spatial effects are applied as specific intercepts with spatial order within the framework. Furthermore a linear term can be incorporated to extend the model as follows:

$$Y = b_0 + b_1 x_1 + \sum_{i=2}^{s} f_i(x_i) + \varepsilon \tag{3.13}$$

The result is semi-parametric models that are particularly convenient for including dummy variables or other contrasts derived from categorical predictors (Lin, 2010: 69).

The method used to estimate ANR is principally the backfitting algorithm. This algorithm iteratively estimates the function f_i in the model (see equations 3.12 and 3.13). This is achieved in the following ways:

i. Initialisation: $b_0 = mean(Y), f_i = f_i^{(0)}, i = 1,2, \dots, s$

ii. Cycle/iterate:

$$f_i = P_i\left(Y - b_0 - \sum_{k \neq i} f_k / x_i\right), i = 1,2, \dots, s$$

iii. Repeat the second step until the individual function, $f_i p$ converge.

The backfitting algorithm calculates the residuals for each of the estimates of $f_i(x_i)$ and

$$Y = b_0 - \sum_{k \neq i} (f_k / x_i)$$

smoothens them against x_i (Lin, 2010: 69). One benefit of using the backfitting algorithm is that it reduces multivariate regression to successive simple bivariate regressions (Bin, 2004: 70).

3.4 M5P Trees

According to Zurada, *et al.* (2011: 357) M5P trees are normal decision trees with linear regression models at the leaves that based its decision on predictions when observations have reached the leaves. It was originally formulated as model tree by Quinlan (1992) as a technique for dealing with continuous class learning problems. Model trees incorporate a conventional decision tree with linear regression functions at the leaves but Wang & Witten (1997: 4) observe that handling enumerated attributes and missing values are not clearly defined in Quinlan's idea. This led to their proposition for an improvement to earlier work particularly as it concerns real-world datasets. In their modification and clarification of how this might effectively be utilised with real-world data, M5P trees was introduced. M5P trees' regression algorithm was developed by Wang & Witten (1997) to process data different from model tree, essentially in the selection of attributes. It is an improvement to the classical model (M5) trees (Quinlan, 1992) in that M5P utilises attributes that predict or forecast results different from the normal theoretic metrics that the classical model tree utilises.

In the study of Holmes, Hall & Frank (1999) this algorithm was to generate rules from datasets. Holmes, *et al.* (1999: 7–8) reported that M5P trees apply the divide-and-conquer method to create sequence for its numeric predictions that is accentuated through a one rule off reading approach. The effective implementation of M5P trees is a three-pronged approach namely splitting the initial tree, pruning the tree and smoothing the tree (Wang & Witten, 1997: 2–3). The first approach requires a splitting criterion which treats standard deviation as a measure of error at each node of the class values. By testing attributes of each node, important attributes that have potential of maximising error reduction is selected. The equation used in calculating standard deviation reduction (SDR) is given as:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} X sd(T) \quad (3.14)$$

where T denotes a set of training instances that reach the node with a set of attributes used for every training case, T_i are subclasses that result after separating the instances that touch the node in line with the selected attribute and sd is the class value standard deviation that reach a node as a measure of the error at that node (Wang & Witten, 1997: 4).

The second approach is pruning the likely error that might ensue from each node of the test dataset. In this stage, a distinction between estimated and actual values is averaged for all the training instances that touch the node. Though this sometime falls below the expected error for unexplained cases, they are compensated by multiplying a factor (P') as follows:

$$P' = \frac{(n+v)}{(n-v)} \quad (3.15)$$

where n is used to denote the size of training instances that touch the node and v represents the total number of model parameters that signifies the class value at that node.

Finally, the third approach in building M5P trees is the smoothing procedure that compensate for severe breaks or interruptions that often happen between nearby leaves at the node of the pruned trees. The smoothing process is given as:

$$P' = \frac{np+km}{n+k} \quad (3.16)$$

in which p' is used to represent the predicted value that moves up to the upper node, p denotes the predicted value that moves to the lower node, m is the model predicted value at this node, n is the total number of training instances that reach the lower node, and k is a constant which usually has a value of 15. The purpose of smoothing is to enhance accuracy of predictions. M5P trees represent its output by a tree structure that distinguishes between all nodes and the leaves (Acciani, *et al*. 2011: 29).

3.5 Artificial Neural Network Models

According to Engelbrecht (2007: 5) the human brain is a complex, nonlinear and parallel computer with the ability to handle tasks at a greater speed such as pattern recognition, perception and motor control than any computer, irrespective of the shortness of time or moments at which event occur for neural systems. The ability of the human brain to learn, memorise and generalise well created research interest to artificially mimic the biological neural systems (BNSs) commonly known as the ANNs. The BNSs are nerve cells also referred to as neurons containing the cell body, dendrite and an axon massively interconnected. The connection between the axon of the neuron and the dendrite of another neuron is commonly referred to as a synapse (see Figure 3.1). The signals are conveyed from the dendrites through the cell body to the axon from where the signals are transmitted to all connected dendrites.

The artificial neuron is modelled after the biological neuron of the human brain. The artificial neurons (ANs) are units in the ANNs that receive one or more inputs as numerical values associated with their respective weights. A bias or threshold level is added as an additional input value to the summation function. The summed value is passed to the next phase to execute the activation function which produces the output from the neuron as shown in Figure 3.2. The detail processes of the ANNs are given in the next paragraph.

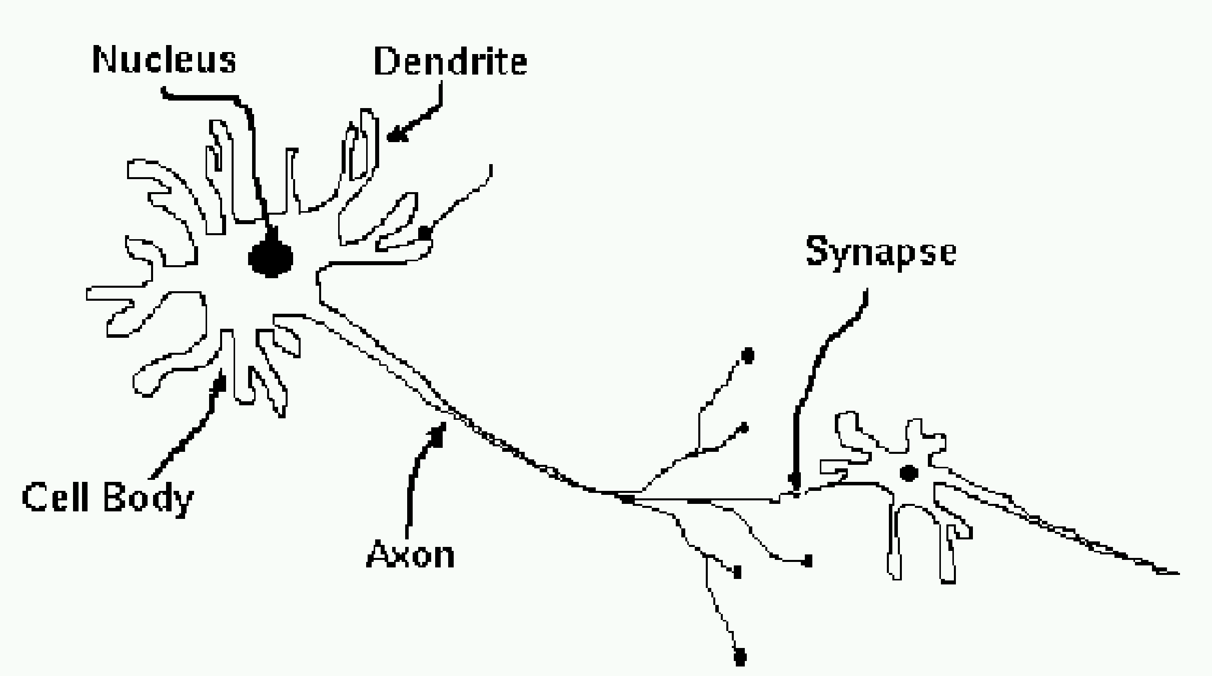

Figure 3.1 A biological neurons (adopted from Engelbrecht (2007: 6)).

The ANNs has multilayer perceptron as the most popular network architecture, consisting of an input layer, an output layer and at least a hidden layer for processing the nonlinear elements. The least (one) number of hidden layers is a reminiscence of the suggestion given in Masters (1993) that one hidden layer should be the initial choice for any practical ANNs design. In this regard, the study of Lin & Mohan (2011: 234) revealed that a single hidden layer is adequate for ANNs to approximate any complex nonlinear function and achieve accuracy (Hornik, 1991). In this study a hidden layer was used to construct the architecture of the ANNs. The values of the input variables are sent into the network via the input layer to the hidden layer. The number of input variables in the network is determined by the configuration of the input data. The output layer is fully connected to the neurons of the input layer and this line of connection runs through the entire units of the network.

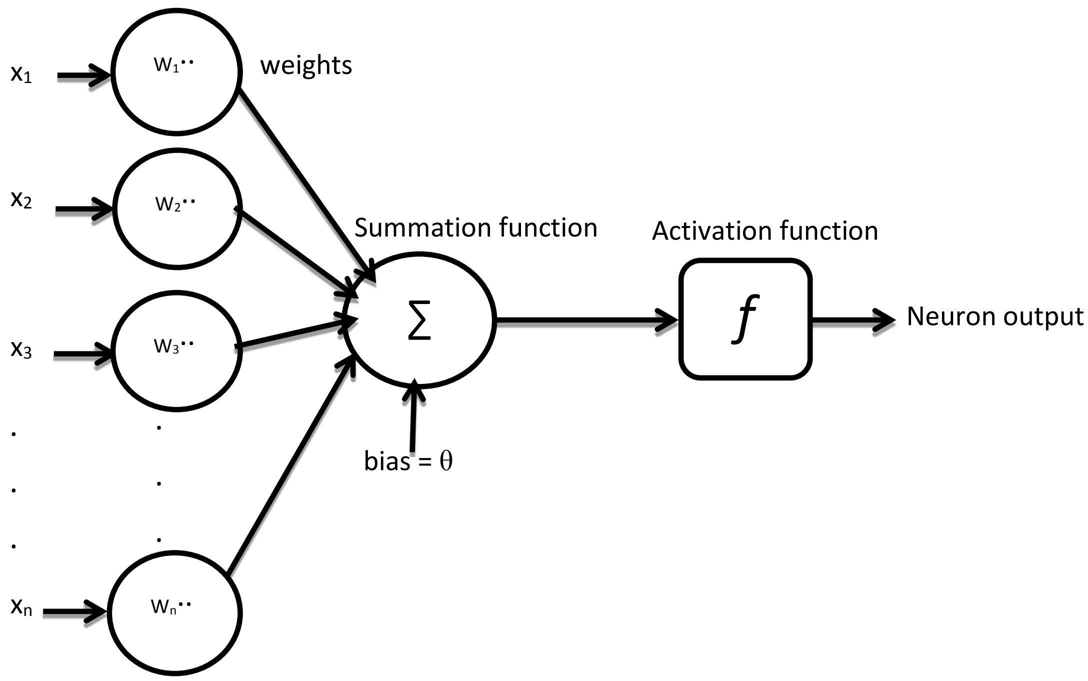

Figure 3.2 The structure of the artificial neuron.

The number of neurons in the hidden layers is a matter of user discretion through a trial and error process. Kwok & Yeung (1997: 631) proposed that the method of finding optimal number of hidden neurons is to begin with a lesser number of neurons, the number should be increased until a desired result is found. The same procedure of optimal selection of the number of hidden neurons as used by Lin & Mohan (2011: 234) and McCluskey *et al*. (2013: 250) was employed in this study. Another important element of the neural networks is the transfer function that determines the relationship between inputs and output (target) of the neuron and its network. The preferred transfer function in this study was the tan-sigmoid used in the neurons of the hidden layer and the linear transfer function was used in the neurons of the output layers. The network compares the output with the actual value to ascertain its accuracy by the total mean squared error. In building a model, the usual practice is to split the data into two or three such as training, testing and validation. The next few paragraphs will dwell on network training.

3.5.1 Artificial Neural Networks Training Algorithms

Training process is a fundamental aspect of ANNs, because failure or success of ANNs is among other factors (such as network architecture, initialisation, parameter values, data pre-processing) dependent upon the training process. The purpose of the training phase is to reduce a cost function usually defined as sum squared error (SSE), mean squared error (MSE) or root mean squared error (RMSE) between the actual and predicted property sale values which is achieved through a series of adjustments to the network weights and biases. ANNs utilise a number of training algorithms including conjugate gradient (CG), LM, and BP being the most widely used, amongst many others.

3.5.1.1 Back Propagation Algorithm (BP)

The BP algorithm is not the first training rule for ANNs (Engelbrecht, 2007: 24). However, in most real estate mass appraisal assessment the BP is the most commonly used algorithm in training the ANNs. The sum squared error given in equation (3.17) is required to evaluate the training process.

$$E(y,w)=\frac{1}{2}\sum_{m}^{M}\sum_{l}^{L}\left(d_{lm}-o_{lm}\right)^{2} \tag{3.17}$$

where y and w are the network input and weight vectors, m is the index of patterns, from 1 to M, in which M denotes the number of training patterns; l is the index of outputs, from 1 to L, in which L is the total number of outputs; k and p are indices of weights, from 1 to N, where N is the total number of network weights and d_{lm} *and* o_{lm} are desired and actual values of the m^{th} output and the l^{th} patterns. The BP utilises the first-order derivative of total error function to find the minimum in error space. Yu & Wilamowski (2011) noted that the gradient g is defined as the first-order derivative of the total error function (equation 3.17) as

$$g = \frac{\partial E(y,w)}{\partial w} = \begin{bmatrix} \frac{\partial E}{\partial w_1} & \frac{\partial E}{\partial w_2} & \cdots & \frac{\partial E}{\partial w_N} \end{bmatrix}^T \qquad (3.18)$$

with the definition of g in equation 3.18, the update rule of the gradient/steepest descent algorithm is written as:

$$w_{r+1} = w_r - \alpha g_r \qquad (3.19)$$

where r is the index of iterations and α is the learning constant (step size taken in the negative direction of the gradient). The training process of the gradient descent algorithm is asymptotic (approaching a curve arbitrarily closely) convergence. The convergent behaviour of BP is dependent on the choice of the initial values of the network connection weights which is related to the network parameters including the learning rate and momentum. This procedure could sometimes be problematic because of the tendency to over train the network and slow convergence. Training speed could significantly be increased by the use of second-order algorithms including Newton and LM, scaled conjugate gradient and Powell-Beale amongst others. Additionally, Openshaw (1998: 1864) suggests that another way of easing this problem might be to use meta-heuristics (GA, PSO etc.) to design and train ANNs. The use of these training algorithms is amongst others what this study is set to investigate.

3.5.1.2 Levenberg-Marquardt Algorithm (LM)

LM was principally designed by Levenberg (1944) and Marquardt (1963) to provide a numerical solution to nonlinear least squares problems. This algorithm has the properties of gradient descent stability (slight variation in training error even if there is a change in training data point) and Gauss-Newton (GN) speed which gives it advantage over all others. These attributes are what carved a unique niche to LM as a versatile ANNs training algorithm because in many respect convergence speed is assured despite complexities of the error surface. According to Yu & Wilamowski (2011a) the central idea of LM is that its training process is combined, switching between two algorithms. Therefore, if the combination coefficient μ is small, GN algorithm is utilised but when combination coefficient is large, the gradient/steepest descent algorithm is employed. Consequently, an adjustment is made to μ parameters so that the network converges to an optimum desired position. The GN algorithm simplifies the calculation of the second-order derivative while the LM is considered a trust-region that modifies the GN method (Battiti 1992: 159).

The sum squared error (E) in equation (3.17) is used to evaluate the training process in LM. The update rule of LM is given as (Yu & Wilamowski, 2011: 465b).

$$\Delta w_r = \left(H_r + \mu I\right)^{-1} g_r \tag{3.20}$$

where I is the identity matrix, g is the gradient vector, the Hessian matrix is given as H and μ is the combination coefficient. The Hessian matrix (H) and gradient vector (g) are defined in equations (3.21) and (3.22) as

$$H = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_1 \partial w_N} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_2 \partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 E}{\partial w_N \partial w_1} & \frac{\partial^2 E}{\partial w_N \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_N^2} \end{bmatrix} \tag{3.21}$$

$$g = \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \\ \cdots \\ \frac{\partial E}{\partial w_N} \end{bmatrix} \tag{3.22}$$

According to Yu & Wilamowski (2011: 466b) the second order derivative of E in equation (3.21), has to be computed in order to perform the update rule in equation (3.20). The process according to Battiti (1992: 148), Yu & Wilamowski (2011a) and Yu & Wilamowski (2011: 466b) is complicated. Furthermore, in the Hagan & Menhaj (1994: 990) and Yu & Wilamowski (2011: 466b) implementation of the LM algorithm, the Jacobian matrix (J) was introduced so that the process of computation will be simplified by avoiding the second-order derivatives as

$$J = \begin{bmatrix} \frac{\partial e_{11}}{\partial w_1} & \frac{\partial e_{11}}{\partial w_2} & \cdots & \frac{\partial e_{11}}{\partial w_N} \\ \frac{\partial e_{12}}{\partial w_1} & \frac{\partial e_{12}}{\partial w_2} & \cdots & \frac{\partial e_{12}}{\partial w_N} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial e_{1L}}{\partial w_1} & \frac{\partial e_{1L}}{\partial w_2} & \cdots & \frac{\partial e_{1L}}{\partial w_N} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial e_{M1}}{\partial w_1} & \frac{\partial e_{M1}}{\partial w_2} & \cdots & \frac{\partial e_{M1}}{\partial w_N} \\ \frac{\partial e_{M2}}{\partial w_1} & \frac{\partial e_{M2}}{\partial w_2} & \cdots & \frac{\partial e_{M2}}{\partial w_N} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial e_{LM}}{\partial w_1} & \frac{\partial e_{LM}}{\partial w_2} & \cdots & \frac{\partial e_{LM}}{\partial w_N} \end{bmatrix} \tag{3.23}$$

By combining equations (3.17) and (3.22), the gradient vector elements can be calculated as

$$\frac{\partial E}{\partial w_k} = \sum_{m=1}^{M}\sum_{l=1}^{L}\left(\frac{\partial d_{lm} - o_{lm}}{\partial w_k} d_{lm} - o_{lm}\right) \tag{3.24}$$

So that the relationship between Jacobian matrix (J) and gradient vector (g) can be presented by

$$g = J^T e \tag{3.25}$$

By combining equations (3.17) and (3.21), the Hessian matrix (H) can be computed as

$$\frac{\partial^2 E}{\partial w_k \partial w_p} = \sum_{m=1}^{M}\sum_{l=1}^{L}\left(\frac{\partial d_{lm} - o_{lm}}{\partial w_k}\frac{\partial d_{lm} - o_{lm}}{\partial w_k} + \frac{\partial d_{lm} - o_{lm}}{\partial w_k \partial w_p} d_{lm} - o_{lm}\right)$$

$$= \sum_{m=1}^{M}\sum_{l=1}^{L}\frac{\partial d_{lm} - o_{lm}}{\partial w_k}\frac{\partial d_{lm} - o_{lm}}{\partial w_k} \tag{3.26}$$

So that the relationship between Jacobian matrix (J) and Hessian matrix (H) can be described by equation (3.27) as

$$H = J^T J = Q \tag{3.27}$$

where matrix Q is the approximated Hessian Matrix (*H*) known as quasi Hessian Matrix. Therefore the LM update rule is achieved by integrating equations (3.20) and (3.24) with equation (3.27).

$$\Delta w_r = (J_r^T J_r + \mu \mathrm{I})^{-1} J_r^T e_r \tag{3.28}$$

The error vector is (e). To implement the LM algorithm, equation (3.28) is used. The procedure is to first calculate Jacobian matrix (*J*), then carryout matrix multiplications of equations (3.25) and (3.27) for more weight update. In the Jacobian matrix (*J*) defined in (3.23), there are $M \times L \times N$ elements that needed to be stored. Wilamowski, Kaynak, Iplikci & Efe (2001: 1779) and Yu & Wilamowski (2011: 467b) report that the LM algorithm is effective for solving problems with small and medium sized training patterns but ineffective in handling large sized training patterns because of its memory limitation. To support this view, Wilamowski & Yu (2010: 931) observe its limitation in solving parity-16 problem of 65,536 patterns and suggested improvement of not storing Jacobian matrix but replacing Jacobian matrix multiplication with vector operations so that problems with unlimited number of training patterns can be solved. The size of training patterns in this study is considered adequate (small) for the LM algorithm, however, for a large size pattern the approach suggested and used in Wilamowski & Yu (2010) should be the preferred.

LM has been used in many fields for ANNs training with optimal results. The only mass appraisal/valuation study found that involved LM is the study of El Hamzaoui & Perez (2011) undertaken in Casablanca, Morocco. The study is of exploratory nature and found the algorithm as a useful model in ANNs training. This research is considered to be more comprehensive in the analysis of residential properties to ascertain its capability in ANNs training.

3.5.1.3 Conjugate Gradient Algorithm (CG)

Unlike BP algorithm that adjust weights in the direction of the negative gradient of the error surface, conjugate gradient algorithm performs searches in the conjugate paths which usually gives convergence that is faster than searches made along gradient descent directions. The algorithm "trades off the simplicity of gradient descent and fast quadratic convergence of Newton's method" (Engelbrecht, 2007: 45). Many variants to conjugate gradient algorithm used in training feed-forward ANNs including Fletcher-Reeves CG, Polak-Ribiére CG, Powell-Beale Restarts CG and Scaled Conjugate Gradient exists. In this study, Powell-Beale Restarts and scaled conjugate gradient are the primary focus. This is because the Powell restart method gives an opportunity of restarting if there is little orthogonality (shift or change) left between the current and previous gradient (Sharma, Sharma & Kasana, 2007: 1116). Again, to save time which is akin to CG during line search at each iteration step, the scaled conjugate gradient utilises a step size scaling approach for line search per iteration.

3.5.1.3.1 Scaled Conjugate Gradient Algorithm (SCG)

Scaled gradient algorithm is another approach used in estimating the step size (Møller, 1993: 529). This algorithm is unique in training feed forward ANNs because it combines LM with conjugate gradient algorithms. Accordingly, the process involves introduction of a scalar parameter denoted as $\beta_r$in CG, where r represents 0,1,2,3…,N, the step size α_r is positive and the path p_r are generated by the equation

$$p_{r+1} = \theta_{r+1} g_{r+1} + \beta_r p_r \tag{3.29}$$

This is the search path, also known as scaled conjugate gradient algorithm. In equation (3.29), parameters that are to be determined are θ_{r+1} and β_r. Thus if θ_{r+1} = 1, the result is a classical conjugate gradient algorithm based on the value of scalar parameters β_r (Andrei, 2007: 403). Conversely, if scalar parameter is 0 ($\beta_r = 0$), then another class of algorithm that supports the choice of parameter θ_{r+1} evolved. Looking at $\beta_r = 0$, it is possible for θ_{r+1} to assume a positive scalar or negative definite matrix. These possibilities are either a change in the paths of gradient descent or Newton algorithms.

If $\theta_{r+1} = 1$, a gradient descent algorithm ensued. But if $\theta_{r+1} = \nabla^2 f(x_r +)^{-1}$, or estimation of it (Andrei, 2007: 403), then it shifts to the path of Newton or Quasi-Newton algorithms. Again if $\theta_{r+1} \neq 1$ is chosen in a quasi-Newton way and $\beta_r \neq 0$, equation (3.29) characterises a blend of quasi-Newton and conjugate gradient techniques. According to Andrei (2011: 325) if θ_{r+1} is a matrix that has information about inverse Hessian of function *f*, then it is better to use $p_{r+1} = -\theta_{r+1} g_{r+1}$ because adding the term $\beta_r p_r$ might affect the path p_{r+1} from being a descent path except if the line search is accurately sufficient.

Scaled conjugate descent algorithm has been applied in ANNs training for parity problems using 20 different initial weight vectors with a momentum set at 0.9 for all simulations. The result indicates that SCG converges faster than standard BP, conjugate gradient with line search and Brayden-Fletcher-Goldfarb-Shanno memory-less quasi-Newton algorithms (Møller, 1993: 532). Orozco & García (2003: 354) used SCG to avoid time consuming line search per iteration of other second order CG algorithms in classifying two types of infants cry. Also Ambarish & Saroj (2016: 500) applied the SCG algorithm in the prediction of soil moisture content for the control of farm irrigation in eastern India. Their study found that SCG performed better than Broyden Fletcher Goldfarb Shanno (BFGS) in training feed-forward ANNs. With this in mind, SCG is proposed for ANNs training in mass appraisal of properties.

3.5.1.3.2 Powell-Beale Conjugate Gradient Algorithm (PBCG)

In conjugate gradient algorithm the search path is reset to the negative position of the gradient periodically. When number of weights and biases in the network equates the number of epochs, the standard reset point is said to occur, however, other reset approaches such as Powell-Beale are used to influence the efficiency of network training. These approaches are proposed by Powell (1977) based on the earlier work of Beale (1972). Powell's proposition is that a restart should ensue if the orthogonality between the new and the former gradient is left to a very minimal amount. This is defined by the following condition to reset the steepest descent path

$$\left|g_{r-1}^{T} g_{r}\right| \geq \alpha \left\|g_{r}\right\|^{2} \tag{3.30}$$

where α is the restart/reset factor which has a range of 0.1, 0.2, … 0.9 (Powell, 1977: 251), g is the gradient and subscript and r is the gradient index. The major shortcoming of this algorithm is the expensive nature of computation of line search requirements for each iteration step. This algorithm has been utilised in ANNs training for identifying the presence of red palm weevil (Al-Saqer & Hassan, 2011: 362). Specifically their study shows that although Powell-Beale conjugate gradient training algorithm consumes less training time, SCG algorithm outperformed Powell-Beale CG in their assessment.

3.5.2 Genetic Algorithm

According to Melanie (1996: 2) the idea to use evolution as an optimisation tool for engineering problems began in the 1950's and the 1960's when several computer scientists including Fraser (1957a); Fraser (1957b); Bremermann (1967) and Reed, Toombs & Barricelli (1967) undertook a study on evolutionary systems. Genetic algorithm (GA) as it is known currently was first described by John Holland in the 1960's and further popularised by Holland in 1975 as a parallel adaptive search algorithm which patterns its work after the theory of biological evolution. GA is used as an optimisation exploration scheme to choose optimal or close optimal network architecture and fitness evaluation function. The processes commence with a population of solutions that represents potential attribute weights. These are evaluated for the purpose of optimal selection of solutions that have a greater chance of producing "offspring" used to form a completely new generation.

In the new population, some members undergo modification through crossover and mutation to establish novel solutions. The process of crossover involves the joining of chromosomes from two parents with similar traits to form two offspring that are alike by swapping similar bits of the parents. This is represented in the following seven-dimensional vectors: – i_1, j_1, k_1, l_1, m_1, n_1, o_1 and i_2, j_2, k_2, l_2, m_2, n_2, o_2. However after crossing the chromosomes, the following offspring are produced: - i_1, j_1, k_1, l_2, m_2, n_2, o_2 and i_2, j_2, k_2, l_1, m_1, n_1, o_1. The changes that occur reveal that there is an exchange of information between possible solutions in crossover. In mutation, the genes of selected chromosomes are altered arbitrary with a possible chance of identical occurrence. The evolutions of different solutions continue until the algorithm attains a global optimum (best) solution, representing the optimal weights (McCluskey & Anand, 1999: 229).

The search for attribute weights is linked to the ANNs in this study. Accordingly, the initial bit of chromosome signifies the weights and biases in the hidden layer. González & Formoso (2006: 22) report that the chromosomes are encoded in binary or real strings. The location in the chromosome keeping the gene has two possible alleles (gene variants that occur on the same locus of the chromosome) represented as 0 and 1. Melanie (1996: 7) report that the GA processes populations of chromosomes successively exchanging a population with another in the search space. The performance of chromosomes is evaluated in each generation based on the fitness function, with the fittest chromosomes to have higher chances of survival. The fitness function as used by Tian & Noore (2005: 46) is defined as:

$$Fitness = \frac{1}{1+err} \tag{3.31}$$

$$err = \sum_{i=1}^{q} \frac{\left(\hat{z}_i - z_i\right)^2}{q} \tag{3.32}$$

where q is the number of property samples used in training, $\hat{z}_i$ and z_i are predicted and actual outputs during BP learning and *err* is the mean squared error (MSE) after training epochs.

3.5.2.1 The hybridisation of BP and GA algorithms

According to Gupta & Sexton (1999: 680) the search operations of the GA towards attaining the global optimum solution does not require differentiable objective function like the BP. The error surface for simple approximation is quite complex with a number of local minima, which suggests that the GA is suitable for searches involving so many minima. The GA searches from one population to the other, paying particular attention to the location of the best solution so far attain while continuously sampling the entire search space. The limitations of GA are that, it is slow because of the random initialisation of genes; the exploration mechanisms employed, and do not always guarantee optimal convergence (Bertini, De Felice & Pizzuti, 2010: 167). The combination of BP and GA might enhance their capabilities and offset the noticeable weaknesses.

The procedures involve a direct amalgamation of the BP and GA with the GA used to optimise weights and threshold values in BP as a basis for modelling property prices. In doing this, the weights of the BP are encoded into a number of chromosomes of the initial population of the GA and other randomly generated chromosomes. The GA having the suboptimal weight set in its candidate solutions tries to further enhance it by searching for more optimal network weights (Papakostas, Boutalis, Samartzidis, Karras & Mertzios, 2005:

172). The early stopping criterion is used to terminate training once learning counts reach the determined values. The next stage is to enter the evolutionary cycles where the fitness function of the GA is evaluated as used in Liang (2008: 205). The fitness value is what describes the effectiveness of the candidate solution. The best solution is one that produces minimum mean squared error or root mean squared error in the training and testing datasets. In this study binary code is used for each gene typifying a single attribute weight. There is equality in the number of independent variables of property and the number of genes.

The combined BP and GA algorithms have demonstrated improved results over the standalone BP, GA and other algorithms used to compare its performance in previous studies. Specifically, the hybrid model has been used to optimise and train ANNs in the prediction of property prices (McCluskey & Anand, 1999), prediction of software cumulative failure time (Tian & Noore, 2005), prediction of future failure data for repairable systems (Liang, 2008), and prediction of river water quality (Ding, Cai, Sun & Chen, 2014) amongst others.

3.5.3 Particle swarm optimisation

PSO is a stochastic population-based search algorithm that can be used for optimisation of any continuous valued problems (Engelbrecht, 2007: 49). This optimisation algorithm was developed by Kennedy & Eberhart (1995) inspired after the social behaviour of birds and fish in their flocking or schooling. PSO belongs to the family of swarm intelligence (simulating animal behaviour in real world) algorithms. Other members of swarm intelligence include genetic algorithm, ant colony optimization, bee colony algorithm, differential evolution and fish-swarm algorithm. The algorithm is related with evolutionary algorithms such as GA's in the sense that they are both population based search algorithms (Hassan, Cohanim, de Weck & Venter, 2004: 18). A considerable number of studies have found it to be very effective in weight optimisation and training of ANNs (Kennedy & Eberhart, 1995: 1947; Zhang, Zhang, Lok & Lyu, 2007: 1036; Suresh, Harish & Radhika, 2015: 273). According to Osman, Omar & Mustafa (2013: 18) this heuristic algorithm is popular amongst academia and industry participants because of its intuitiveness, ease of application and flexibility in solving complex problems, but its application in mass appraisal of properties is doubtful. This lack of empirical analysis involving PSO as an optimiser of ANNs in the mass appraisal industry necessitated its investigation in this study. Its optimisation procedure begins with initialising a population of random solution to seek for optimality through update of generations. In its search for possible solutions commonly referred to as particles, PSO goes through problem space for current and best particles. Zhang, *et al.* (2007: 1028) notes that the birds that flock together through searching space are called "particles". Thus these "particles" soar high with a known velocity to find optimal position after a series of iterations.

At various intervals of iterations, the velocity is adjusted by each particle according to the level of its momentum and impact of its best position (P_b) which also influences its neighbour's best position (P_g). These two positions commonly referred to as best values, are (1) P_b which is the best fitness that a particle could reach at the moment while, (2) P_g is the position of the best fitness or solution that particle's neighbour has so far reached. Modelling the social thinking and behaviour of birds is predicated on the understanding that when birds leave a particular position (region) in the search space they tend to return to the original region where successes have been achieved. There are two update parameters, the position (m) of birds and their velocity (v) as they scour the search space for the best solution. These two (position and velocity) are however influenced by the best values of the particle and that of its neighbours. Furthermore, the position and velocity of a particle is updated in the following equation:

$$v_{il}(t+1) = wv_{il}(t) + c_1 r_1 [P_{bil}(t) - m_{il}(t)] + c_2 r_2 [P_{gil}(t) - m_{il}(t)] \tag{3.33}$$

$$m_{il}(t+1) = m_{il}(t) + v_{il}(t+1) \tag{3.34}$$

in which $v_{il}(t+1)$ is used to denote the i^{th} particle's velocity at the l^{th} iteration, $m_{il}(t+1)$ denotes the i^{th} particle's position at the l^{th} iteration. Furthermore w is the weight inertia, designed to control the influence of the previous velocities, c_1 and c_2 are cognitive and social factors that influence the movement of every particle against the two best positions (distinct versus global social experience) and r_1 and r_2 denotes random values given between 0 and 1.

3.5.3.1 Proposed hybrid algorithm

PSO improved its performance through weights (w) adjustment and parameter settings. This has the ability to search global space for optimal weights but its disadvantage is slow speed in its global search. However, the BP algorithm has the ability to search local space for weights but its limitation lies in its weakness for global search for weights. Bringing these two algorithms together will complement each other. PSO is used at inception for global search of optimal weights through acceleration of training speed. Nevertheless, if there is a change in value from the predefined number or if the fitness functions remain unchanged for some generations, the gradient descent searching process is activated. BP is activated to search around global optimum to enable the hybrid model to attain optimality more quickly. In this study PSO is combined with the BP algorithm for prediction of market values of properties. The main motivation of building the hybrid model is to have a combination of social thinking ability (*gbest*) found in PSO with the local search feature of BP.

According to Mirjalili, Mohd, & Sardroudi (2012: 11126) the first part of equation (3.29), $wv_{il}(t)$, gives PSO exploratory ability in searching the weights space, while the second and third parts, $c_1r_1[P_{bil}(t) - m_{il}(t)]$ and $c_2r_2\left[P_{gil}(t) - m_{il}(t)\right]$, denote private thinking and collaboration of particles, respectively. There is a random initialisation of particles and velocities in the range [0, 1] at the beginning in PSOBP. The particle's fitness value is evaluated after each initialisation to achieve desired positions for P_b and P_g. Furthermore, the best particles (solution) out of the current particles are stored while the quality of a solution is assessed and appraised via the update rule (equation 3.29) so that a new group of particles are formed. If the new particle flies beyond the boundary $[m_{min}, m_{max}]$, the new position will be set at m_{min} or m_{max}, if a new velocity is beyond the boundary $[v_{min}, v_{max}]$, the new velocity will be set at v_{min} or v_{max}. After updating the optimal solution so far attained, each new particle's fitness value is evaluated such that the worst particle is changed by the stored best particle (if the *i*th particle's different position is superior than P_{ib}, then P_{ib} is made the new position for the *i*th particle. If the optimal position of all new particles is superior to P_g, then P_g is updated). The inertia weights *w* is reduce based on the strategy used for selection.

Should the current position of P_g for n generations remain unchanged, the BP algorithm is used to search around P_g for some epochs provided the result is better than P_g. If the outcome is superior to P_g then this is compared with the worst particle out of the current particles. The essence is to replace the worst particle with the best particle. After reduction of inertia weights *w* based on strategy for selection, it will give the global optimum output. The hybrid model PSOBP is used to optimise and train ANNs for classification and functional approximation. The PSO is used to find weights of the feed forward neural network with every particle representing set of weights. There are three encoding approaches and strategies used for every particle including vector encoding, matrix encoding and binary encoding. The binary encoding strategy stem from the proposition of Kennedy & Eberhart (1997: 4104) whereby every particle is seen to move to closer and farther places of the hypercube by flipping different numbers of bits with values zero and one, similar to chromosome in GA. Also the number of changes to the bits per iteration is used to describe the overall velocity of the particle. One feature of PSO that makes it different from GA is its flexibility to use both binary- and real numbers with abilities of avoiding known local optima.

PSO has been used for predictions in different fields. Specifically Garro & Vázquez (2015) used a PSO to design ANNs and to test the fitness function. Three bioinspired algorithms were used namely PSO, second generation of particle swarm optimisation (SGPSO) and a new model of PSO. The study found the model useful in network designs at different scenarios and parameter settings. The study of Suresh, Harish & Radhika (2015) utilises PSO and BP for inpatient length of stay prediction in hospital and found PSO as optimal replacement for predictions with BP. Again the study of Chang & Hsieh (2011) combined PSO with BP algorithms in forecasting exchange rates between USD and NTD, a Taiwanese local currency and found that PSOBP achieved a higher accuracy match with the actual exchange rate than standalone PSO and BP algorithms. Also Lv, Wang, Xie & Wei (2008) used a hybrid PSO and neural network to predict a tobacco pest in Chongqing, China and found the hybrid model to outperform BP algorithm and autoregressive moving average (ARIMA). Again the study of Zhang *et al*. (2007) built hybrid PSOBP and adaptive PSOBP algorithms to train feed forward neural networks in functional approximation and classification problems and found PSOBP to perform better than adaptive PSO, standalone PSO and BP algorithms. The simple reason that could be adduced for these optimal results is because PSO efficiently underwent global search for best model weights during network training.

3.5.4 Cuckoo Search Algorithm

To overcome limitations of standard BP noted earlier, a novel meta-heuristic search algorithm known as CS is utilised to optimise ANNs in this study. CS has been utilised with other training algorithms to optimise network, improve performance and converges to global optimum (Kaveh, Bakhshpoori & Ashoory, 2012:13; Nawi, Khan & Rehman, 2013a: 419–421; Nawi, Khan & Rehman, 2013b: 108–112).

The CS algorithm was developed by Yang & Deb (2009) based on the required brood parasitic action of some cuckoo species of birds that lay their eggs in the nests of other birds that are not of their category. As it is with birds, they may incubate the eggs to maturity if they do not recognise differences between their own eggs with the ones laid by cuckoos. However if the pretender's (cuckoo) egg is discovered by the host bird, two things may occur, to discard the egg or simply abandon the current nest and build another elsewhere. If the eggs are not discovered by the host bird's, because of similarities in colour and pattern between cuckoo and host bird's eggs, it will lessen the chance of cuckoo eggs being left in desolation and thus enhance their chances of living.

There are three basic rules in CS algorithm:

1. Every cuckoo bird lays an egg at a time, and deposits its egg in arbitrarily chosen nest;
2. The optimal nests with high calibre of eggs will transcend over to the next generations;
3. The number of accessible host nests is unchanged, and the chances of a host bird discovering an egg that are alien to its own which might result in either abandoning the nest or throwing out the eggs is given by a probability $p_a \in [0,1]$.

For the maximisation problem, the standard or fitness of a solution can only be measured to the value of the objective function. In CS algorithm, each egg in the nest portrays a solution, say *x*, while the cuckoo egg portrays a new solution as $x^{(i+1)}$. The aim is to use the new and possibly improved solutions (cuckoo) to replace a not-so-improved solution in the nests. The three rules set out initially are utilised here in setting basic steps for CS algorithm, summarised as pseudo codes below:

Start

Step 1: Objective function f(x). $x=(x_1, x_2,\ldots, x_d)^r$

Step 2: Generate initial population of *n* host nest *i* = 1,2,3,…,n

Step 3: *While* (*t* min < *MaxGeneration*) or (*stop criterion*)

Step 4: *G**o*** Get a Cuckoo randomly by Lévy flights and assess it fitness/suitability F_i

Step 5: Select at random a nest *j* among *n*

Step 6: If $F_i > F_j$ *Then*

Step 7: Change *j* by the new solution, *End If*

Step 8: A portion (p_a) of worst nests are jettisoned and new ones are built

Step 9: Retain the best solutions (or nests with quality solutions)

Step 10: Grade the solutions and find the current best, *End While*

Step 11: Post-process results and visualisation.

End

Lévy flight is a straight flight route changed by a sudden 90^{o} turn, leading to an irregular scale-free search pattern. Nawi, Khan & Rehman (2013a: 416) reported that the flight movement of many animals and insects follows a random pattern. It is a random walk whereby the step-lengths have a chance dispersal that is purposefully tailed. The Lévy flight process is an integral part of a CS that is used for both global and local search (Walton, Hassan, Morgan & Brown, 2011: 712). While generating novel solutions $x^{(i+1)}$, for a cuckoo *i* a Lévy flight process is performed as:

$$x_i^{(t+1)} = x_i^t + \alpha \oplus L\acute{e}vy\ (\lambda), \tag{3.35}$$

Where $\alpha > 0$ = the step size which is linked to scales of problem requiring solutions. The notation $\alpha = 1$ is utilised. The equation is crucial in the stochastic random walk equation. It is an effective procedure in exploration of the weight space for the best solution having a longer step-length. The product $\oplus$ signifies entry wise multiplication functions. Essentially, the Lévy flight gives a random walk while the random step length is deduced from a Lévy distribution.

$$Lévy \sim u = t^{-\lambda}, \quad (1 < \lambda \leq 3), \tag{3.36}$$

The procedure forms a random walk process with a power law step length transfer with a heavy tail. Most of the new solutions are generated by Lévy random walk after efficient search around the optimum solution. Further to this, a reasonable portion of these new solutions that are generated should be randomly chosen in distant fields from the current optimum solution. This is to avoid entrapment of the network into local minimum. CS has been applied as an optimisation algorithm for various assignments such as finding optimal features/paths for the neural network; optimising the network parameters for engineering designs (Yang & Deb, 2010); health sector (Valian, Mohanna & Tavakoli, 2011); short-term electricity price forecasting (Taherian *et al*, 2013) amongst others.

3.5.4.1 Hybridisation of CS with LM and BP Algorithms

The proposed model works under three rules as noted earlier: a). Preference for best source by safeguarding the quality nests or solutions; b). exchange of host eggs with new solutions or cuckoo eggs generated randomly through Lévy flights; and c). detection of few other cuckoo eggs by host birds and substituting based on the quality of local random walks (Yang & Deb, 2009). The cycle consists of several initialisation steps of best nest or solution, the number of host nests available is unchanged, and the cuckoo egg is uncovered by the host bird with a probability, p [0,1].

In this search algorithm, every good nest or solution signifies an optimal solution in the weight space and biases utilised for optimisation of ANNs. The CS is used to determine network structure and find the best weights and biases. Yi, Xu & Chen (2014: 2) report that the BP network is the first part, determined based on the number of input and output parameters and the size of each cuckoo. The next part concerns the application of CS technique to optimise weights and biases of the BP and LM networks. The CS technique is used to implement the initialisation, determine the fitness function, update and select the position of the operator, replace and eliminate operator so that good cuckoo individual(s) with best fitness are establish. The process of optimisation continues if there are no immediate best weights and biases for cuckoo to select from, until cuckoo attains the last cycle/epoch in finding optimality from global space. One notable character of cuckoo birds is when it is near optimum weights, a slow movement ensued (Lévy flights) until best weights is selected. Furthermore, selected best weights are passed to BP or LM for network training. The last epoch is usually the point at which the network attains a minimum level of mean squared error (MSE) or root mean squared error (RSME). Moreover, the neglected solution by the cuckoo is replaced with a new best nest.

The flow diagram for CS hybridisation with LM and BP algorithms used in this study is given below:

Start

Step 1	Dataset for training and testing input
Step 2	Initialisation of the network with cuckoo search
Step 3	Best weights and biases are passed to Back Propagation/Levenberg Marquardt
Step 4	Feed Forward Neural Network utilises weights for assessment
Step 5	Network calculates and corrects error backward
Step 6	Cycle continue until optimum is attained

Step 7	Training of CSBP and CSLM is complete, while mean squared error < stopping criteria
Step 8	If yes, then stop and store best weights
End	

3.6 Support vector machines (SVMs)

The SVMs is a relatively new class of techniques developed by Vapnik for solving classification and regression problems (Vapnik, 1999: 988 and Zurada *et al*., 2011: 361). It became prominent due to its attractive features and its successful application that traversed several fields. SVMs uses input data and classify it into one of two groups or classes. To achieve effectiveness, SVMs first utilise a set of training input datasets, map them into multidimensional space, and use regression to discover a hyperplane that is suitable for separating the two class inputs. In selecting the best hyperplane for data classification, the one that characterises the largest separation or margin between the two classes is the most preferred. Generally, the bigger the margin, the smaller the classifier's generalisation error will ensue.

Furthermore finding the margin requires a construction of two separate hyperplanes on either side of the canonical plane (a canonical plane is an n-dimensional hyperplane used in separating a data point) that is "pushed up against" the two datasets (Amarappa & Sathyanarayana, 2014: 436). According to Cui & Curry (2005: 609) the underlying philosophy behind the margin classification is similar to OLS but the problem is stated and solved as a nonparametric optimisation problem and not as a parametric (maximum likelihood) problem. In pricing of property with SVMs, for example, a sample drawn from a known distribution $P(x, y)$ in which y takes on two values. For linearly separable instance, the decision rules defined by an optimal hyperplane separating the binary decision classes is given in equation (3.37) in terms of the support vectors (Shin, Lee & Kim, 2005: 130).

$$Y = sign\left(\sum_{r+1}^{N} y_r \alpha_r (x \cdot x_r) + c\right) \tag{3.37}$$

The vector $x = (x_1, x_2, x_3,\ldots, x_n)$ corresponds to an input and the vectors x_r, where $r = 1, 2,\ldots N$, are support vectors, Y denotes the outcome, y_r denotes the class value of the training example x_r, c and α_r are factors that determine the hyperplane, while the $\cdot$ symbolise the inner product. However, for a non-linearly separable instance, a kernel function K is introduced in equation (3.37) to generate the inner products which construct machines that have different types of nonlinear decision surfaces in the input space as

$$Y = sign\left(\sum_{r+1}^{N} y_r \alpha_r K(x \cdot x_r) + c\right) \tag{3.18}$$

There exist a number of support vector machines that are used in practice. The study of Shin, Lee & Kim, (2005: 130) reported three of the many different types of SVMs used in constructing the decision rules. These are: (1) polynomial kernel function, depicted as $K(x \cdot x_r) = (x \cdot x_r + 1)^d$, where d is the degree of the polynomial kernel; (2) a radial basis kernel function, depicted as $K(x \cdot x_r) = \exp(-1/\delta^2(x - x_r)^2)$ and (3) a two-layer neural network machine with kernel function, given as $K(x \cdot x_r) = S[(x \cdot x_r)] = 1/[1 + exp\{v(x \cdot x_r) - j\}]$, where v and j are parameters of the sigmoid $S[(x \cdot x_r)]$ that satisfy the inequality $j \geq v$.

This thesis utilises polynomial and radial basis kernel functions; and added the normalised polynomial kernel function to assess the price of properties. By utilising the different type of machines (kernel functions), the algorithm can construct many learning machines so that performance can be evaluated. According to Cui & Curry (2005: 604) there is no metatheory designed to guide researchers and practitioners in the choice of a particular kernel function. This leave the expert with the option of selection based on domain knowledge or preference supplemented by numerical results in the choice of a kernel function (Boser, Guyon & Vapnik, 1992). Kernel function is used to successfully map the original input space into a high dimension space. To achieve high accuracy the parameters of the kernel function must be tuned. Two parameters namely the C bound and γ kernel parameter must be determined, however, the parameters are varied to select the optimal values for the best performance (Lam *et al.*, 2009: 222). The study of Tay & Cao (2001: 313–316) reports on the sensitivity of SVMs to the parameters setting, of which, a chosen values could either lead to over-fit or under-fit of the training data. The implication of this is that optimal result could only be achieved after several runs. According to Tay & Cao, the C bound has a range of between 1 and 100. In this study the limiting value of C was set at the minimum value of 1.0 and thereafter adjusted slightly upward while the kernel parameter γ was equally adjusted once a fixed value was set for C to minimise error.

SVMs initially suffer a setback with its training, particularly when large datasets are used for quadratic programming (QP) solver to train. With this limitation, Osuna, Freund & Girosi (1997: 16) noted that since the problem with QP is the requirement for enormous large scale data, the breaking down of the problem into a series of smaller sized QP problems will better optimise SVMs. In keeping with Osuna *et al's* idea, examples must be added and subtracted to keep the matrix size constant. Therefore Platt (1998: 5) observes this process as ineffective, because it will cause the training example in all numerical QP optimisation steps to obey Karush-Kuhn-Tucker (KKT) conditions, irrespective of the strategy employed. Details on KKT conditions for QP problems could be found in Gill, Murray & Wright (1981). In all strategies deployed in Osuna's proposition, a numerical QP solver is used which according to Platt (1998: 5) is extremely complicated because of the myriads number of numerical accuracy concerns that is required. In finding a simple algorithm that can optimise and train SVMs, Platt (1998: 6) introduced sequential minimal optimisation (SMO) to swiftly tackle the SVMs QP problem without adding additional matrix space and numerical QP optimisation phases at all. To guarantee convergence, SMO disintegrates all QP problems into sub-problems without using QP solver to provide a solution.

The process of finding new optimal values requires that SMO selects two combined Lagrange Multipliers (LM) to find optimal values for SVM updates. Thus the benefits of using SMO for SVMs training is that it avoids the numerical QP optimisation to solve for two LMs analytically

3.7 Chapter summary and conclusion

The methodologies based on their underlying philosophies, strengths, weaknesses and improvements made to enhance their capabilities were reviewed in this chapter. The HRMs have been the most widely used models in the mass appraisal environment. The benefit of using it lies in the transparency of the process but its limitation is in the parametric restrictions and rigidity. Consequently a number of non- or semi-parametric techniques, including the artificial neural networks, support vector machines, additive nonparametric regression, M5P trees, GWR, SEM and SLM have been introduced into the mass appraisal environment. Also the review paid particular attention to the training algorithms of the ANNs. The most widely used algorithm is the back propagation, but this is said to suffer from certain limitations which have been remediated in other fields and are proposed in this study.

4. CHAPTER FOUR

4.0 Data and Modelling Procedures

4.1 Introduction

Although gathering of property data for the administration of taxes or other purposes is the responsibility of assessors, the quality of data for effective assessment of property prices is a matter of great significance to the modeller. According to IAAO (2013: 5) standard on mass appraisal of real property, the accuracy of values is dependent on the completeness and precision of property characteristics and market data. More so, Borst (2007: 134) opine that the conclusion reached in empirical analysis of multivariate models are linked to the data used in their development. The purpose of this chapter is to explain the procedures leading to the assessment of property prices within the confines of the various techniques chosen for this study. Consequently the sources, nature and quality of data, including identification and removal of unwanted transactions that would impede effective assessment are highlighted. Also detail description of the variability among the selected property attributes, the accuracy test statistics and IAAO benchmark tests used are amongst others discussed.

4.1.1 The Cape Town assessment

The city of Cape Town is 944 square miles; it is located on a peninsula beneath the Table Mountain on the southwest coast of South Africa. It is the provincial capital of the Western Cape and the legislative capital of the country. There are other municipalities that are adjacent to it including Swartland and West Coast to the north; drakenstein, Cape Winelands and Stellenbosch to the north-east; and Theewaterskloof, Overberg and Overstrand to the south-east. The city of Cape Town has centralised its property tax assessment under the city valuation office (CVO). The CVO is the body responsible for the valuation of 915,148 residential and commercial properties in Cape Town (KPMG, 2015: 12). The CVO has created 1555 neighbourhoods and 31 submarkets for the purpose of assessment and reassessment.

4.1.2 The data

The CVO Cape Town, South Africa granted access to the database of 3526 properties. The original data included 46 property variables and features which were examined to ensure their suitability for the assessment. The process of data cleaning begins with reduction in the number of property variables. The number of variables was reduced to 11 (before the inclusion of the *x*, *y* coordinates) because some of them are cost related or information about them is repeated by other variables, adding them will increase the correlation amongst variables (Table 4.1), resulting in unacceptable levels of multicollinearity. There are other variables that describe the property such as single-family dwelling, property number and street name depicted with a STRAP code. STRAP code is a means by which a property is easily identified and located in South Africa. The neighbourhood codes in the dataset were used to create submarkets and thus serve a better purpose than the STRAP code; hence this is excluded from the list of variables. Furthermore, the variable number of living area was excluded because of a high level of correlation to number of bedrooms. Other variables removed including Jacuzzi, detached sauna, squash and tennis courts which have infrequent number of occurrences.

Table 4.1 Property variables

Variable	Description
Assessed_value	Assessed market value of properties in Rand (A Rand is USD15)
Num_beds	Total number of bedrooms
Quality	Quality grade of construction
Condition	Physical condition of property
Storey	Total number of storeys
Bld_style	Building architecture style & design
View	Quality grade factor of property
RMOS	Reverse month of sale
Property size	Size of property in square metres
Pool	Size of swimming pool in square metres
Submkt	Locational variable identifying submarkets

The second stage was the conversion of text (non-numeric) data into numeric values (Table 4.2). It was also observed from the original file that most text data have similar descriptions. These were renamed so that assessment would not be affected (see columns 2 and 4 of Table 4.2). The problem of having variables with similar name(s) is inability of a model to differentiate between the contributions of the specific variables thereby increasing the standard error of some coefficients (Mark & Goldberg, 1988: 109). Furthermore an initial examination of the descriptive statistics on the selected variables was carried out on the original data file. This becomes necessary to ascertain the suitability of data for the analysis. Suitability of data for analysis is predicated on its freedom from biases that might result from missing values, outliers and multicolinearity. These, if not properly mitigated, could lead to assessment error. A careful scrutiny of the initial descriptive statistics (Table 4.3) revealed the need for data cleaning. Accordingly, the number of transactions shows that not all data have complete information (see for instance number of bedrooms, quality, condition, building style and property view). Again the observation revealed that the number of bedrooms has an extreme and unrealistic value (33) in a transaction. The assessed value range from zero to R61200000 transaction and the average value was R4519609.56. The property size contains zero transactions and a very high standard deviation (81.843). The largest property was 717 square metres and the average single family dwelling was 176 square metres. Another area of concern was the very low unrealistic transaction for swimming pool (-15).

Table 4.2 Conversion of text data to numeric values

Variable name	Text data (dummy)	Numeric format	Rename
Quality	Poor	1	Q_poor
	Fair	2	Q_fair
	Average	3	Q_average
	Good	4	Q_good
	Very good	5	Q_vgood
	Excellent	6	Q_excellent
Condition	Poor	1	C_poor
	Fair	2	C_fair
	Average	3	C_average
	Good	4	C_good
	Excellent	5	C_excellent
Building style	Sub-economic	1	S_subeco
	Unconventional	2	S_unconv
	Conditional	3	S_conven
	Georgian victor	4	S_georgvic
	Cape Dutch	5	S_ capedutch
	Maisonette	6	S_maisonette
	Mediterranean	7	S_medit/t
	Group housing	8	S_grouph
View	Partially obstructed	1	V_partiallyob
	Below average	2	V_belowave
	Average	3	V_average
	Above average	4	V_a/average
	Panoramic	5	V_panoramic
	Excellent	6	V_excellent

The third phase of cleaning requires elimination of unrealistic transactions from the data. Consequently transactions with more than ten bedrooms were removed from analysis. The remaining transactions have the number of bedrooms ranging from zero to 10. Three transactions were found to contain zero bedrooms. These were recoded as "1" and added to one bedroom category. Again the variable building style has a dummy variable "group housing" (style_group) in one transaction that was added to style_medit/t category.

Transactions written as "NA" were identified and removed in some variables including quality and condition. The "null" transactions in the number of storeys were excluded. Transactions that have above 4 storeys were included in the modelling activity but recoded as 3 storeys because of the few number of occurences. Transactions with zero storeys were considered ground floor or bungalow. The two transactions found in the category were recoded to reflect one storey category. Again transactions with missing values in property size and condition were removed from data. The "office property" transaction that appears in building style was removed from analysis. The recoding is consistent with the study of Guan *et al.* (2008: 403–406) to ameliorate the problem of dimensionality which is capable of reducing the strength of a model.

Table 4.3 Initial descriptive statistics of variables

Variable name	No. of transaction	Minimum	Maximum	Mean	Std. deviation
Assessed_val	3526	0	61200000	4519609.56	3408977.935
Num_bed	3525	0	33	3.57	1.119
Quality	3522	1.0	6.0	3.505	0.6284
Condition	3524	1.0	5.0	3.520	0.6343
Storey	3522	0	7.0	1.53	0.580
Bld_style	3524	1.0	7.0	3.030	0.4238
View	3525	1.0	6.0	3.582	0.9635
Property size	3526	0	717	176.40	81.843
Pool	3526	-15	154	13.85	18.187

Also, all transactions having occurrence between 1 and 11 square metres and greater than 150 square metres for swimming pool were excluded from analysis. For the assessed values, all transaction with zero prices and those of less than R250,000.00 and greater than R40,000,000.00 were excluded from the analysis. For the property size all transactions greater than 600 square metres and those of between 1 and 30 square metres were excluded. The number of transactions after data cleaning used for analysis was 3232 (see Table 4.4 for the final descriptive statistics). This sample was considered adequate for the analysis because the methods themselves are not in any case controlled by the number of properties that can be handled (McCluskey & Anand, 1999: 231).

Depending on the method, continuous, categorical and binary variables were used in this study. For BP-ANNs, GWR, and the hybrid systems, variables were left in their categorical form. But when linear, semi-log, log-log regression, ANR, SEM, SLM, M5P trees and SVMs were used for assessment; the binary dummy format was used. For example, variables such as property view and quality were coded between 1 and 6 when assessed in BP-ANNs, GWR, and the hybrid systems but when linear, semi-log, log-log regression, ANR, SEM, SLM, M5P trees and SVMs were used this was coded 1 and 0 (with 1 if the categorical condition is met, 0 if the condition is not met) leaving out the most commonly occurring category to avoid the dummy variable trap that might lead to a state of perfect multicollinearity and resulting in the failure of the regression program (Greene, 2003: 118; Borst, 2007: 50).

The number of variables (11) was carefully chosen using pragmatic and realistic approaches. They are also considered likely to be value significant in line with previous multivariate analysis of this nature. Table 4.4 summarises the final descriptive statistics of the cleansed data used in this study. This illustrates the variability within the cleaned data. The mean assessed value in the sample data is R4,483,474.29 while the average property size is 177 square meters. The smallest property assessed is 31 square metres. The average number of storey in the sample is two. The average property has four bedrooms. The total transactions were stratified into 70% training and 30% testing datasets in accordance with acceptable norm that data should be partitioned to allow for modelling or training and testing. The stratification was done using WEKA explorer via a resample procedure.

Table 4.4 Final descriptive statistics of variables

Variable name	Mean	SD	Minimum	Maximum
Assessed_value	4483474.29	3117754.039	824000	38000000
Beds	3.56	0.992	1.0	10
Quality	3.494	0.6155	1.0	6.0
Condition	3.506	0.6269	1.0	5.0
Storey	1.52	0.553	1.0	3.0
Bld_style	3.034	0.4300	1.0	7.0
View	3.581	0.9630	1.0	6.0
RMOS	14.885	8.1645	1.0	29.0
Size	177.45	78.905	31	599
Pool	13.97	18.362	0.0	154

4.1.3 Spatial consideration

Location plays an important role in the achievement of accuracy. The study of McCluskey & Borst (2007: 313) reported that a number of methods have been used in specifying and calibrating the effect of location. These methods are market segmentation, a neighbourhood delineation variable, accessibility measures, explicit use of location and advanced model specification methods. A detailed discussion on these methods is provided by McCluskey & Borst (2007). The methods are sometimes used in combination with others in a single study depending on the model capability, specification and availability of location element(s) within the data. In this study, the neighbourhood delineation variable and explicit use of location (*x*, *y* coordinates) were used to specify the effect of location as variables within the techniques that are compatible with explicit location coordinates. The delineated neighbourhoods and submarkets defined by Tax Assessors for ease of identification were used. Bourassa *et al*. (2010: 141) defined submarkets as geographic regions or non-contiguous clusters of dwellings with comparable features and/or hedonic prices. Accordingly a total of 181 neighbourhoods were extracted from the property dataset and used to establish their submarkets. In all there are 31 submarkets in the city of Cape Town as noted earlier.

The study of Tu, Sun & Yu (2007: 391) reported that property units are grouped into a cluster particularly if the market is continuous. However the property market of Cape Town is discontinuous as the study of Tu *et al*. (2007: 391) noted concerning the characteristic of the whole property market. There are noticeable physical features such as green parks, road and mountains that were used in dividing the market. The study of Bourassa *et al*. (2007: 146) noted that because of similarities in the prices of property characteristics in a submarket, the likelihood of errors to be correlated within a submarket is higher than across submarkets. They suggested that controlling the submarkets would diminish assessment error and improve prediction of property prices. There are a number of ways by which this could be achieved but a simpler way of dealing with assessment errors is the assignment of dummies to the identified submarkets, model property prices for each submarket and make adjustment to errors of predicted results within the submarket (Bourassa *et al.*, 2007: 146). In effect Wilhelmsson (2002: 96) used 13 submarket dummy variables from previously defined administrative parish to account for location in the HRMs. Bourassa *et al*. (2007: 151) created 33 submarket dummies to improve their results. Also Lin & Mohan (2011: 231–233) created 66 dummies from neighbourhood codes to form submarkets/clusters in their analysis. In this study out of the 31 submarkets, the 181 neighbourhoods that contain the sampled property dataset were found in 15 submarkets (submarkets codes: Submkt48, Submkt50, Submkt52, Submkt53, Submkt54, Submkt55, Submkt56, Submkt64, Submkt65, Submkt66, Submkt67, Submkt68, Submkt69, Submkt70, and Submkt73). The base submarket in this assessment is Submarket 54. To assess the likely spread of sales transactions across the selected submarkets, the frequency test was carried out in Table 4.5.

Table 4.5 Description of properties sold across the submarkets

Submarket code	No. of Neighbourhoods with sales in a submarket	% of sales	Frequency of sales in a submarket	% of sales
Submkt48	1	0.54	41	1.266
Submkt50	5	2.70	227	7.024
Submkt52	16	8.65	560	17.33
Submkt53	18	9.73	405	12.53
Submkt54	16	8.65	686	21.23
Submkt55	6	3.24	246	7.611
Submkt56	9	5.41	266	8.230
Submkt64	3	2.16	12	0.371
Submkt65	2	1.08	02	0.062
Submkt66	15	8.11	110	3.303
Submkt67	57	31.4	346	10.71
Submkt68	18	10.3	220	6.807
Submkt69	6	3.24	38	1.176
Submkt70	8	4.32	70	2.166
Submkt73	1	0.54	03	0.093
15	181	100.00	3232	100.00
SD		7.381		6.532
Average		6.671		
Min		0.54		0.062
Max		31.4		21.23

The ratio between the total number of submarkets and the submarkets used in this study is 48%. This is a marginal representation of the entire submarkets in the city of Cape Town. Furthermore, the frequency count reveals that the standard deviation of observations in each submarket is 7.38% with an average coverage of 6.67%. This depicts that a reasonable spread across the 15 submarkets were achieved in this analysis.

4.1.4 Temporal consideration

Having dealt with the spatial effect to account for influence of location on property price, next is to incorporate a time trend (temporal effect) into the models. The need to incorporate the time trend in CAMA has assumed a very significant place because of increase in property values in some geographical areas and decrease in property values in other areas. There are a number of methods that are used in practice to adjust for time including paired sales analysis, resale analysis, incorporating time variable in regression models and using current appraised value (Gloudemans, 1990: 84). Recently, the studies of Borst (2008: 34–35) and Borst (2009: 29–30) suggest the use of reverse month of sale (RMOS), Fourier expansion, quarterly or half yearly binary (dummy) variables and a linear spline method for incorporating time trend into a model.

The dummy scale variable is used to describe the period a property was sold. Borst (2009: 30) and Borst (2014: 130) report that the approach involves creating a set of binary (1, 0) variables for the whole period of sales. The studies use a form of notation to depict each quarter, for instance, QD1 represents a quarterly dummy that equals 1 for RMOS = 1–3 for the first quarter of sales and 0 otherwise; QD2 dummy equals 1 for the second quarter of sale, RMOS = 4–6 and 0 otherwise. Suppose there are 10 quarters, the ninth quarter is computed as QD9 dummy equals 1, RMOS = 25–27 and 0 otherwise. Consequently, this study used the reverse month of sale and quarterly or half yearly binary (dummy) variable to account for the influence of time on CAMA. The two were used independently for different models. The semi-annual binary dummy variable was used in the modelling activity of linear, semi-log, log-log models, SEM, SLM, ANR, M5P trees and SVMs, while the reverse month of sale was used for BP-ANNs, GWR and the hybrid systems. Therefore, in using the reverse month of sale, the most recent month in the sample is assigned RMOS = 1; the next month is assigned RMOS = 2; and this continues to the oldest month which is assigned RMOS = 29 because the sales period spans from January 2012 to May 2014. Five semi-annual dummies were created in this study. Accordingly, the notation SA1 was used to represent the first half year that equals 1 for RMOS = 1–6 otherwise 0; SA2 was used to depict the second half year that equals 1 for RMOS = 7–12, otherwise 0. For the third half year (SA3) that equals 1 for RMOS = 13–18, otherwise 0. The process continue until the last quarter, but to avoid dimensionality and dummy trap problems that would reduce the strength of the models and multicollinearity, the most occurring category was excluded from the analysis. Table 4.6 gives a summary of the specific types of transformation of time trend variables used in the study.

Table 4.6 Illustration of time trend variables used in the study

Year and month of sale	RMOS	SA1	SA2	SA3	SA4
May–14	1	1	0	0	0
Apr–14	2	1	0	0	0
Mar–14	3	1	0	0	0
Feb–14	4	1	0	0	0
Jan–14	5	1	0	0	0
Dec–13	6	0	1	0	0
Nov–13	7	0	1	0	0
Oct–13	8	0	1	0	0
Sep–13	9	0	1	0	0
Aug–13	10	0	1	0	0
Jul–13	11	0	1	0	0

The table contains the two variables used to illustrate time trends used in this study and a few of the months of sale are also represented.

4.2 Performance measurement

The GWR4.09, *GeoDa*, RapidMiner Studio 7.4, Spatial Analysis in Macroecology (SAM v4.0); MatlabR2013b, Statistical Package for the Social Science (SPSS) v21 and Waikato Environment for Knowledge Analysis (WEKA, 3.6) explorer were used in this analysis. The root mean squared error (RMSE), squared correlation coefficient (R^2), and mean absolute error (MAE) were used to explain the model's predictive capabilities. RMSE, MAE, and R^2 are given in the equations:

$$RMSE = \sqrt{\sum_i \left(y_i - \hat{y}_i\right)^2 / n} \tag{4.1}$$

The RMSE defined above is the square root of the average of the squared values of the estimation error. RMSE usually assign higher weights for large errors than smaller errors. It is one of the extensively used tools for accuracy measurement.

$$MAE = \frac{\sum_{i=1}^{n} \left|\hat{y}_i - y_i\right|}{n} \tag{4.2}$$

The MAE in this study is expressed in rand; it deals with errors uniformly based on their sizes. It is the average of the absolute values of the estimation errors.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{4.3}$$

Thus apart from squared coefficient correlation R^2 which measures correlation between the actual and predicted values that must have a reasonable goodness of fit well above 50%, the model with the lowest MAE and RMSE is considered a better technique. Again another important measure of model performance used is the Akaike's Information Criterion (AIC). This accuracy test statistic is designed to choose from many competing models the optimal model based on the maximum likelihood criterion of the model parameter. McCluskey & Borst (2011: 304) report that the estimates of β_i are based on the least squares and the maximum likelihood estimates of the model parameters are identical. This permit the expression of AIC in terms of statistics available from the OLS regression as:

$$AIC = n + n\,\mathrm{In}(2\pi) + n\,log\left(\frac{RSS}{n}\right) + 2K \tag{4.4}$$

where K is the estimated parameters in the model that include the intercept and $\acute{\sigma}^2$ (McCluskey & Borst, 2011: 304) while the RSS denotes the sample residual sum of squares. The model with the minimum AIC is adjudged the best in this study.

Furthermore to measure the model's quality in line with consistency and uniformity, the benchmark test acceptable to the International Association of Assessing Officers (IAAO, 2013) was applied. These are coefficient of dispersion (COD) and price related differential (PRD). The COD measures assessment uniformity and variability of how much the value ratios vary from the median ratio.

$$COD = \frac{\sum \left| Z_i - Z_m \right|}{nZ} X100 \qquad (4.5)$$

PRD on the other hand is used to measure consistency of valuation ratios between low valued properties and high valued properties. In essence this technique measures the vertical equity of appraisals which should be less than 1.0 (progressivity), because measures above 1.0 suggest regressivity (IAAO, 2013: 29).

$$PRD = \frac{\dfrac{\sum Z_i}{n}}{\dfrac{\sum PV_i}{\sum AV_i}} \qquad (4.6)$$

where, Z_i is used to depict PV_i/AV_i; and PVi is the predicted value of the ith property; AV_i is the assessed value; Z_i is the ratio between the predicted value and assessed value of the ith property and Z_m is the median of Z_1.

The benefit of utilising performance measures is that it evaluates predictive precision of models. Additionally the choice of a number of performance measures in this study represents a complete attempt to calculate and compare the performance of different approaches in the context of mass appraisal assessment.

4.3 Chapter summary and conclusion

This chapter set the stage for a comprehensive modelling of property prices with the different methods. The original data of 3526 transactions was reduced to 3232 after the removal of transactions with missing, extremes and unrealistic values. The large number of variables was also reduced to 11 following pragmatic approaches to avoid the problem of multicollinearity and dimensionality. The text data were coded to enhance suitability for modelling. Furthermore, to enhance the modelling activity, the neighbourhoods identified via codes created by tax assessors were used to form submarkets in order to account for the spatial aspect of the data. Five bands were used to create dummy variables which represent semi-annual notations that account for the temporal aspect within the data. The RMOS was also used but calibrated for models that do not require excessive use of the dummy variables. The chapter also describes the techniques used for evaluation of model performance.

5. CHAPTER FIVE

5.0 Analyses and Discussion of Results

5.1 Introduction

This chapter contains the empirical part of the study. Section 5.2 presents the results for the three hedonic regression models (OLS, semi-log and log-log) to establish a baseline model for the Cape Town property market. The baseline model is used for comparison to the proposed- and all other models discussed. However, this is preceded by a test for neglected nonlinearities in the OLS regression model (McCluskey *et al.*, 2012: 281). Section 5.3 presents the results of spatial models (GWR, SEM and SLM) designed to address the spatial effects of the hedonic regression. Section 5.4 presents the results of other mass appraisal models namely SVMs, ANR and M5P trees. Section 5.5 is the main thrust of this thesis; the ANNs modelling trained with BP, LM, PBCG and SCG algorithms. The hybrid systems (GABP and PSOBP) which select attribute weights from global and local spaces used in prediction of property prices are presented, and lastly, the chapter conclude with detailed comparison of the performance of techniques evaluated.

As noted previously, several adjustments were made to the data to allow for an easy comparison to be made between all models. For instance, to generalise findings, the 100% data used for ANNs, SVMs, ANR, M5P trees and the hybrid systems was preferred relative to the training and testing sets designed to avoid excessive training and overfitting. The stratification of data became necessary because of the sensitivity of ANN based models to overfitting (Gonzalez, Soibelman & Formoso, 2005: 316). Consequently, the 100% data allow for ease of comparison with the 100% data used for linear, semi-log, log-log and the spatial models.

5.2 Establishing a baseline model for the Cape Town property market

The study first undertakes the test for neglected nonlinearities in the OLS to establish the case for using the nonlinear techniques.

5.2.1 Test for neglected nonlinearities

Peterson & Flanagan (2009: 158) report that those supporting the use of ANNs based their argument on the neglected nonlinearities observed in the linear models. Consequently, a RESET test that applied the predictions of the ANNs as applied by Peterson & Flanagan (2009); and McCluskey *et al.* (2012) was used to justify the application of nonlinear models. The predictions of ANNs are included as an additional regressor in the OLS. Therefore if the OLS specification is:

$$y = X\beta + \mu \tag{5.1}$$

and the predictions of the ANNs are contained in the vector m, then a test of neglected nonlinearities is equivalent to a t-test on HO: $\varphi = 0$ in the regression:

$$\mu = X\beta + \varphi m + v \tag{5.2}$$

The null-hypotheses of no neglected nonlinearities is easily rejected at any conventional level of significance which reveals the need for using the nonlinear techniques in the data. Table 5.1 provides the summary.

Table 5.1 OLS test for neglected nonlinearities

Variable	Coefficients	Std. coefficient	VIF	Std. error	T	P Value
Constant	184128.64	0	0	108996.235	1.689	.091
ANNs_Pred	.169	.302	5.434	.020	8.279	.000
Beds1	-268912.33	-.016	1.029	268324.101	-1.002	.316
Beds2	59294.49	.012	1.158	86323.676	.687	.492
Beds4	-15482.94	-.005	1.350	55025.949	-.281	.778
Beds5	84891.11	.018	1.294	84908.868	1.000	.317
Beds6	-41474.57	-.005	1.193	140961.806	-.294	.769
Beds7	-199621.72	-.011	1.069	300670.729	-.664	.507
Beds8	-718658.19	-.023	1.048	489621.563	-1.468	.142
Beds9	-563217.54	-.010	1.014	900162.925	-.626	.532
Beds10	-1611303.76	-.020	1.023	1278560.173	-1.260	.208
Q_Poor	281408.02	.008	1.898	779474.055	.361	.718
Q_Fair	282547.11	.018	1.268	270141.889	1.046	.296
Q_Good	120315.50	.042	2.572	71753.480	1.677	.094
Q_Vgood	692168.80	.049	1.145	236679.916	2.924	.003
Q_Exec	224460.42	.018	1.391	226330.294	.992	.321
C_Poor	671187.54	.022	1.901	659403.960	1.018	.309
C_Fair	10129.23	.001	1.265	213404.801	.047	.962
C_Good	-21277.10	-.007	2.507	71271.145	-.299	.765
C_Excell	51451.23	.008	1.524	124717.101	.413	.680
Storey_2	1740.85	.001	1.500	54616.097	.032	.975
Storey_3	180262.75	.021	1.369	158106.703	1.140	.254
S_Subecon	-5557.41	.000	1.020	737259.479	-.008	.994
S_Unconven	-23656.01	-.003	1.262	157940.060	-.150	.881
S_Georvicto	208584.53	.016	1.035	202106.906	1.032	.302
S_Capedutch	-304658.87	-.014	1.019	354544.091	-.859	.390
S_Maisonett	108745.01	.008	1.042	213405.683	.510	.610
S_Medt	-40112.35	-.001	1.034	575263.505	-.070	.944
V_Part/obs	-50059.37	-.007	1.091	109684.505	-.456	.648
V_Bel/av	-32586.80	-.002	1.019	293559.725	-.111	.912
V_Ab/ave	153332.36	.051	1.283	53134.430	2.886	.004
V_Panor	343931.27	.085	1.510	77445.606	4.441	.000
V_Excell	-11128.77	-.001	1.127	185853.749	-.060	.952
Submkt48	-256405.29	-.020	1.158	213843.225	-1.199	.231
Submkt50	-173579.83	-.031	1.492	106282.625	-1.633	.103
Submkt52	-245923.90	-.065	1.821	79270.687	-3.102	.002
Submkt53	-88541.59	-.021	1.565	84031.206	-1.054	.292
Submkt55	519394.25	.097	2.317	127637.310	4.069	.000
Submkt56	158524.51	.031	1.529	100060.322	1.584	.113
Submkt64	-291337.73	-.012	1.034	371839.580	-.784	.433

Submkt65	-43415.82	-.001	1.008	897615.767	-.048	.961
Submkt66	-150383.35	-.019	1.183	133362.619	-1.128	.260
Submkt67	151260.73	.033	1.574	90234.646	1.676	.094
Submkt68	243594.95	.043	1.422	105282.361	2.314	.021
Submkt69	-110396.79	-.008	1.156	221757.391	-.498	.619
Submkt70	-275218.52	-.028	1.303	174352.198	-1.579	.115
Submkt73	-196250.27	-.004	1.011	734343.680	-.267	.789
SA1	22488.97	.006	1.440	69159.491	.325	.745
SA2	-75546.55	-.021	1.466	67391.449	-1.121	.262
SA3	-60514.74	-.017	1.456	68269.379	-.886	.375
SA5	1074.55	.000	1.424	69578.385	.015	.988
Size	2.778	.000	2.155	413.752	.007	.995
Pool	-599.422	-.008	1.316	1389.581	-.431	.666
Dependent variable = Residual		AIC = 100046.796				

5.2.2 The baseline regression model

The three regression models (linear, semi-log and log-log) are tested based on their R^2, adjusted R^2, AIC and F-statistics to reveal levels of their statistical acceptability and confidence. Apart from the RESET test undertaken above, the study of McCluskey (2016: 130) notes that the log transformation makes it possible to explore the existence of nonlinear relationship in the data. In the process of conversion a constant k was introduced and added to log (x) to avoid error generated by log transformation.

Table 5.2 Goodness-of-fit measurements for linear, semi-log and log-log models

Model	Regression		
	Linear	Semi-log	Log-log
R^2	0.589	0.694	0.687
Adjusted R^2	0.583	0.689	0.682
F-statistic	89.441	141.527	137.045
AIC	103057.028	1461.717	1533.562
Log likelihood	-51474.6	-676.958	-712.881

The coefficient of determination (R^2) shows the explainability of variation in property prices across the three regression models. Accordingly the parameter estimates reveal that 58.9%, 69.4% and 68.7% of the variation in property values are explained by linear, semi-log and log-log models respectively. There was a reduction in the level of explanation to 58.3%, 68.9% and 68.2% when adjustment was made to the general population as revealed in the adjusted R^2 statistic. The adjusted R^2 revealed acceptability of the results since the maximum value is 100%. Again, the three regression models exhibit high F-statistics at $P < 0.01$; linear regression (89.441); semi-log (141.527); and log-log (137.045). In terms of their log likelihood statistics, the linear model has -51474.6, the semi-log and log-log models have -676.958 and -712.881, respectively. In all, the semi-log outperformed the other regression models (linear and log-log). However, the goodness of fit criterion used in Table 5.2 was not sufficient to establish a baseline regression model for the city of Cape Town because the three models have different structure. The COD as used by Borst (2007: 160) can be used for direct comparison of the three models (Table 5.3). Additional measures of model performance including Median Ratio, Mean Ratio, RMSE, MAE and Price Related Differentials are also used. The mean and median ratios are part of the quality assurance benchmark tests recommended for industry use. The standard stipulates that a ratio of between 0.90 and 1.10 meets the required standards.

Table 5.3 Performance comparison of linear, semi-log and log-log models

Test	Regression		
	Linear	Semi-log	Log-log
Median ratio	1.055	1.027	1.032
Mean ratio	1.087	1.044	1.046
PRD	1.08	1.10	1.10
COD	26.7	22.1	22.6
MAE	1234015	1091961	1127104
RMSE	1997911	1951184	1995487

The results in Table 5.3, reveal the semi-log model (1951184) to perform better in comparison to linear and log-log models (1997911 and 1995487) in terms of RMSE. According to Limsombunchai, *et al*. (2004: 196), a model with the lowest RMSE is considered the best in terms of prediction accuracy. The MAE value also reveals the semi-log model (1091961) to predict prices that are closer to the assessed values than the linear and log-log models (1234015 and 1127104). The results show marginal performance of the log-log to linear models in terms of predicting property prices closer to the assessed values. Furthermore, the semi-log model has the best performance in terms of the mean ratio, median ratio and COD while the only slim advantage of the linear model was that it had a slightly better PRD relating to the vertical equity of prices. Therefore on the strength of these results the baseline model for the Cape Town property market is the semi-log model.

The detailed results reveal the regression coefficients, t-statistics and indicators of level of significance of the three regression based models. This is presented in Table 5.4, Table 5.5 and Table 5.6. The test of multicollinearity using the variance inflation factor (VIF) indicates that in all assessments (linear, semi-log and log-log), none of the regression coefficients are inflated.

Table 5.4 Linear regression model coefficients

Variable	Coefficients	Std. coefficient	VIF	Std. error	T	P Value
Constant	1925246.24			158212.73	12.169	.000
Beds1	-95531.07	-.003	1.029	427538.96	-.223	.823
Beds2	15630.32	.001	1.158	137541.94	.114	.910
Beds4	37641.80	.006	1.350	87676.96	.429	.668
Beds5	291784.11	.028	1.293	135278.58	2.157	.031
Beds6	629564.89	.035	1.182	223563.63	2.816	.005
Beds7	637132.87	.016	1.066	478398.83	1.332	.183
Beds8	5007485.25	.075	1.017	768486.58	6.516	.000
Beds9	171405.83	.001	1.013	1434258.10	.120	.905
Beds10	3867852.90	.022	1.015	2029838.16	1.905	.057
Q_Poor	-202524.20	-.003	1.897	1241655.55	-.163	.870
Q_Fair	-61100.14	-.002	1.267	430244.67	-.142	.887
Q_Good	442444.77	.071	2.537	113563.08	3.896	.000
Q_Vgood	2618454.27	.084	1.123	373413.86	7.012	.000
Q_Exec	2756016.95	.102	1.282	346187.56	7.961	.000
C_Poor	363167.64	.005	1.900	1050457.93	.346	.730
C_Fair	-52365.85	-.002	1.264	339991.77	-.154	.878
C_Good	247138.27	.039	2.505	113526.24	2.177	.030
C_Excell	-209527.37	-.015	1.519	198402.30	-1.056	.291
Storey_2	613462.18	.098	1.377	83371.75	7.358	.000
Storey_3	2392759.12	.126	1.264	242111.40	9.883	.000
S_Subecon	280819.94	.003	1.019	1174492.02	.239	.811
S_Unconven	2103625.07	.107	1.143	239491.88	8.784	.000
S_Georvicto	572529.82	.021	1.034	321948.83	1.778	.075
S_Capedutch	-286716.32	-.006	1.018	564814.73	-.508	.612
S_Maisonett	92098.37	.003	1.042	340034.39	.271	.787
S_Medt	-38174.25	.000	1.034	916498.59	-.042	.967
V_Part/obs	176691.38	.012	1.090	174756.93	1.011	.312
V_Bel/av	-139569.28	-.003	1.019	467739.72	-.298	.765
V_Ab/ave	485613.16	.074	1.268	84179.59	5.769	.000
V_Panor	1327260.85	.150	1.382	118050.99	11.243	.000
V_Excell	1994698.18	.081	1.075	289304.63	6.895	.000
Submkt48	-2515927.09	-.090	1.089	330435.18	-7.614	.000
Submkt50	-2094033.00	-.172	1.281	156943.99	-13.343	.000
Submkt52	-1316265.15	-.160	1.603	118522.76	-11.106	.000
Submkt53	-1136953.04	-.121	1.424	127714.31	-8.902	.000
Submkt55	3852948.92	.328	1.460	161459.66	23.863	.000
Submkt56	1391301.39	.123	1.439	154642.04	8.997	.000
Submkt64	-2176342.91	-.042	1.023	589279.11	-3.693	.000
Submkt65	-1892688.68	-.015	1.006	1429077.40	-1.324	.185

Submkt66	-830327.69	-.048	1.166	210958.29	-3.936	.000
Submkt67	-874903.05	-.087	1.463	138579.95	-6.313	.000
Submkt68	-445484.83	-.036	1.373	164824.71	-2.703	.007
Submkt69	-2032349.27	-.070	1.095	343996.14	-5.908	.000
Submkt70	-2293589.64	-.107	1.206	267271.65	-8.581	.000
Submkt73	-1698941.19	-.017	1.008	1168266.11	-1.454	.146
SA1	-35065.35	-.004	1.439	110172.13	-.318	.750
SA2	-187413.09	-.024	1.463	107292.86	-1.747	.081
SA3	-42573.42	-.005	1.455	108746.12	-.391	.695
SA5	132591.65	.016	1.423	110819.77	1.196	.232
Size	9156.69	.232	1.789	600.68	15.244	.000
Pool	7025.48	.041	1.300	2200.61	3.193	.001
Dependent variable: Assessed values						

The regression coefficients reveal the contribution of individual attributes to property price in linear regression. In effect, the result shows that for every increase in property size, there is an increase of R9,156.69 in property price. The base bedroom is a three bedroom single family property. It therefore follows that properties with two, four, five, six, seven, eight and ten bedrooms contribute more than three bedrooms but properties with one and nine bedrooms are worth less than three bedrooms in this study. The base quality grade factor is average quality (q_average). It thus follows that properties of fair quality (q_fair), good quality (q_good), very good quality (q_v/good), and excellent quality (q_excell) grade factor are worth more than poor quality (q_poor) properties. The quality grade factor is very significant to home buyers all over the world and this also to a large extent determines peoples' willingness to pay a named amount for a property. The base for condition of property is average condition (c_average). This reveals that poor condition (c_poor) and good condition (c_good) properties are worth more while properties having fair and excellent conditions are worth less. The base for the storey is storey_1. Properties that are found on two and three storeys (storey_2 and 3) are worth more. The base building style is conventional style (S_con). It thus reveals that sub-economic (S_subeco), unconventional (S_uncon), Georgian victor (S_g/victo), and maisonette (S_maison) building styles are worth more, while Cape Dutch (S_c/dutc) and Mediterranean T (S_medt) building style are worth less. In terms of property view, the base is average view of property (V_avera). The coefficients show that properties that are partially obstructed (V_part/ob), above average (V_a/avera), panoramic (V_panora) and excellent views (V_excell) are worth more while properties whose view is below average (V_b/avera) are worth less. The base submarket is located in submkt54. Apart from properties in submarkets (Submkt55 and Submkt56) that are worth more, all properties in the remaining submarkets are worth less. The base time of sale is in the fourth (SA4) semi-annual notation. Properties that are sold in the fifth semi-annual time of sale (SA5) are worth more by R132,591.65, while properties sold during the first (SA1), second (SA2) and third (SA3) semi-annual time of sale are worth less. Properties with a swimming pool are worth more by R7,025.48 than properties without a swimming pool.

Table 5.5 Semi-log regression model coefficients

Variable	Coefficients	Std. coefficient	VIF	Std. error	T	P Value
Constant	14.722			.024	622.852	.000
Beds1	-.086	-.013	1.029	.064	-1.352	.176
Beds2	-.034	-.018	1.158	.021	-1.675	.094
Beds4	.038	.033	1.350	.013	2.909	.004
Beds5	.071	.039	1.293	.020	3.522	.000
Beds6	.092	.029	1.182	.033	2.754	.006
Beds7	.118	.017	1.066	.071	1.649	.099
Beds8	.524	.045	1.017	.115	4.562	.000
Beds9	.085	.004	1.013	.214	.399	.690
Beds10	.201	.007	1.015	.303	.663	.507
Q_Poor	-.080	-.006	1.897	.185	-.433	.665
Q_Fair	-.002	.000	1.267	.064	-.036	.972
Q_Good	.089	.082	2.537	.017	5.227	.000
Q_Vgood	.266	.050	1.123	.056	4.773	.000
Q_Exec	.350	.075	1.282	.052	6.768	.000
C_Poor	.058	.005	1.900	.157	.371	.711
C_Fair	-.019	-.004	1.264	.051	-.379	.705
C_Good	.048	.044	2.505	.017	2.843	.004

C_Excell	-.066	-.027	1.519	.030	-2.240	.025
Storey_2	.148	.137	1.377	.012	11.904	.000
Storey_3	.342	.104	1.264	.036	9.458	.000
S_Subecon	.070	.004	1.019	.175	.396	.692
S_Unconven	.169	.050	1.143	.036	4.722	.000
S_Georvicto	.117	.024	1.034	.048	2.439	.015
S_Capedutch	-.077	-.009	1.018	.084	-.910	.363
S_Maisonett	-.006	-.001	1.042	.051	-.125	.900
S_Medt	.233	.017	1.034	.137	1.700	.089
V_Part/obs	.100	.039	1.090	.026	3.820	.000
V_Bel/av	-.019	-.003	1.019	.070	-.265	.791
V_Ab/ave	.115	.101	1.268	.013	9.184	.000
V_Panor	.242	.158	1.382	.018	13.720	.000
V_Excell	.287	.068	1.075	.043	6.638	.000
Submkt48	-.914	-.190	1.089	.049	-18.520	.000
Submkt50	-.570	-.270	1.281	.023	-24.299	.000
Submkt52	-.339	-.237	1.603	.018	-19.123	.000
Submkt53	-.248	-.152	1.424	.019	-13.000	.000
Submkt55	.502	.247	1.460	.024	20.800	.000
Submkt56	.202	.103	1.439	.023	8.737	.000
Submkt64	-.810	-.091	1.023	.088	-9.200	.000
Submkt65	-.609	-.028	1.006	.213	-2.851	.004
Submkt66	-.205	-.069	1.166	.032	-6.492	.000
Submkt67	-.220	-.126	1.463	.021	-10.635	.000
Submkt68	-.120	-.056	1.373	.025	-4.874	.000
Submkt69	-.733	-.147	1.095	.051	-14.273	.000
Submkt70	-.829	-.224	1.206	.040	-20.763	.000
Submkt73	-1.024	-.058	1.008	.175	-5.868	.000
SA1	-.014	-.010	1.439	.016	-.845	.398
SA2	-.035	-.026	1.463	.016	-2.157	.031
SA3	-.021	-.016	1.455	.016	-1.310	.190
SA5	.003	.002	1.423	.017	.206	.837
Size	.002	.271	1.789	.000	20.674	.000
Pool	.001	.050	1.300	.000	4.451	.000
Dependent variable: Ln Assessed values						

For the other two models (semi-log and log-log) results show that properties with less than three bedrooms contribute less to property prices. Likewise properties with poor quality grade factor showed a reduction in price. Properties with poor and good conditions are worth more than those of fair and excellent condition. Again properties with more than two storeys are worth more. Property size and swimming pool all have positive correlation on price in the semi-log model. Similar variables have positive influence on property price in the log-log model. The three regression models reveal results that suggest differing significant variables. The model with least significant variables (28) is linear regression; the next is the semi-log regression (34) and the model with the highest number of significant variables (35) is log-log regression. A number of variables have the appropriate signs in semi-log and log-log regression models than in linear regression.

Table 5.6 Log-log regression model coefficients

Variable	Coefficients	Std. coefficient	VIF	Std. error	T	P Value
Constant	13.492			.083	162.123	.000
Beds1	-.076	-.008	1.034	.093	-.811	.417
Beds2	-.032	-.011	1.171	.030	-1.063	.288
Beds4	.056	.034	1.363	.019	2.917	.004
Beds5	.117	.045	1.291	.029	3.989	.000
Beds6	.177	.039	1.169	.048	3.652	.000
Beds7	.216	.021	1.062	.104	2.076	.038
Beds8	.754	.045	1.017	.167	4.504	.000
Beds9	.282	.009	1.010	.312	.903	.367
Beds10	.678	.015	1.010	.441	1.536	.125
Q_Poor	-.120	-.006	1.898	.271	-.445	.656
Q_Fair	.008	.001	1.266	.094	.090	.928
Q_Good	.133	.085	2.540	.025	5.356	.000
Q_Vgood	.380	.049	1.124	.081	4.672	.000
Q_Exec	.527	.078	1.282	.075	6.990	.000
C_Poor	.087	.005	1.900	.229	.379	.705
C_Fair	-.038	-.006	1.264	.074	-.510	.610
C_Good	.066	.042	2.509	.025	2.646	.008
C_Excell	-.100	-.028	1.519	.043	-2.311	.021
Storey_2	.216	.138	1.411	.018	11.739	.000
Storey_3	.500	.106	1.276	.053	9.432	.000
S_Subecon	.128	.005	1.020	.256	.502	.616
S_Unconven	.242	.049	1.144	.052	4.644	.000
S_Georvicto	.162	.023	1.035	.070	2.311	.021
S_Capedutch	-.080	-.006	1.016	.123	-.648	.517
S_Maisonett	.053	.007	1.051	.074	.712	.476
S_Medt	.273	.014	1.035	.200	1.364	.173

V_Part/obs	.130	.035	1.089	.038	3.425	.001
V_Bel/av	-.036	-.004	1.018	.102	-.355	.722
V_Ab/ave	.171	.104	1.268	.018	9.329	.000
V_Panor	.357	.162	1.385	.026	13.856	.000
V_Excell	.437	.071	1.078	.063	6.924	.000
Submkt48	-1.300	-.187	1.094	.072	-18.019	.000
Submkt50	-.813	-.267	1.280	.034	-23.775	.000
Submkt52	-.482	-.235	1.612	.026	-18.624	.000
Submkt53	-.344	-.146	1.424	.028	-12.374	.000
Submkt55	.722	.246	1.460	.035	20.510	.000
Submkt56	.357	.126	1.392	.033	10.768	.000
Submkt64	-1.169	-.091	1.024	.128	-9.097	.000
Submkt65	-.874	-.028	1.006	.311	-2.805	.005
Submkt66	-.297	-.069	1.165	.046	-6.453	.000
Submkt67	-.305	-.121	1.460	.030	-10.114	.000
Submkt68	-.153	-.050	1.371	.036	-4.266	.000
Submkt69	-.961	-.133	1.120	.076	-12.685	.000
Submkt70	-1.117	-.209	1.241	.059	-18.907	.000
Submkt73	-1.452	-.057	1.009	.255	-5.700	.000
SA1	-.026	-.013	1.439	.024	-1.072	.284
SA2	-.056	-.029	1.462	.023	-2.409	.016
SA3	-.033	-.017	1.456	.024	-1.413	.158
SA5	-.001	.000	1.423	.024	-.031	.976
Ln Size	.304	.259	1.899	.016	18.946	.000
Ln Pool	.014	.044	1.318	.004	3.855	.000

Dependent variable: Ln Assessed values

In general the results of the log transformed regression are better than the linear regression, thereby improving the predictive accuracy. The study of McCluskey (2016: 135) observed that it is important to have quality control measures such as the log transformation to verify the results of a predictive model before such estimates are used for taxation. This analysis shows that the semi-log regression models can be used to achieve the objective of tax assessors in fixing taxes for properties in Cape Town. As noted earlier, due to the limitations of the hedonic regression models various improvements were made to enhance its predictive accuracy. The next section of this analysis is concern with the use of the improved versions of the hedonic regression modelling.

5.3 Prediction of property prices with spatially varying and weighted regression models

The spatially varying and weighted regression models are used to predict property prices. Although a number of these models exist, this study utilised the GWR, SEM and SLM techniques. The weight matrix for SEM and SLM is designed in such a way that permits the Euclidian distance to be estimated via the x, y coordinates. Furthermore the coefficients are estimated by the maximum likelihood techniques. The goodness of fit and performance comparison of GWR, SEM and SLM are summarised in Table 5.7. The results represent a robust explanation of the variability in property prices by the models. In comparison, the SEM and SLM have similar results with those of GWR as revealed in their R^2 (0.746, 0.735 and 0.708) with autoregressive models slightly outperforming the GWR. This shows that 74.6%, 73.5% and 70.8% of the variability in the property prices are explained by the autoregressive (SEM and SLM) and non-stationarity (GWR) models.

The R^2 of SEM, SLM and GWR reveal improvement over those of the linear, semi-log and log-log models reflecting the importance of spatial models in remediating the spatial bias limitation of the traditional or global hedonic regression models. Additionally the study of Wilhelmsson (2002: 96) noted that SEM and SLM can correct the shortcoming of location dummy variable in a HRM. The lagged variable in SLM (see Table 5.10) is what enhances the R^2 relative to the regression based models. Again the AIC statistic reveals similarities among the three models but the GWR (102368) produced an AIC goodness of fit that is slightly below SEM (101701) and SLM (101756). However, in terms of the log likelihood, the GWR (101846.7) had a better fit than SEM (-50812.5) and SLM (-50839.2).

Furthermore, results reveal lag coefficients (Lambda (SEM) and Rho (SLM)) for the autoregressive models. In terms of the overall lag coefficients performance the SEM (0.920082) outperform the SLM (0.827761). Both are highly significant ($P < 0.0000$) and usual for observations that exceed 1000 which also reflect the asymptotic nature of the analytical expression used for the variance (Anselin, 2005: 208). The RMSE test statistic reveals the SEM (1570763) to perform better in comparison to the other techniques (GWR (1684620) and SLM (1604163)). The results show a slight outperformance of the SLM over the GWR. The MAE value also reveals the SEM to be nearer in property price prediction term to the actual assessed value than other models (GWR and SLM) with the SLM performing slightly better than the GWR. The median and mean ratios also reveal a better performance of the SEM to the other model.

Table 5.7 Goodness-of-fit and performance comparison of GWR, SEM and SLM

Model	Spatial models		
	GWR	SEM	SLM
R^2	0.708	0.746	0.735
AIC	102368	101701	101756
log-likelihood	101847	-50812.5	-50839.2
Median ratio	1.0454	0.9982	1.0021
Mean ratio	1.0772	1.0178	1.0233
PRD	1.0804	1.0323	1.0412
COD	21.80	16.51	18.42
MAE	1011117	937451.9	1004769
RMSE	1684620	1570763	1604163

The IAAO benchmark test reveals that in terms of their COD and PRD the SEM has the best performance over the GWR and SLM. In all, the analysis reveals the SEM to be the best performing model compared to the GWR and SLM.

The detailed results of the spatial models are summarised in Table 5.8, Table 5.9 and Table 5.10. The GWR results reflect the importance of localised spatial influences within the Cape Town property market (Table 5.8). The GWR parameter estimates vary at each of the 3232 observation points is revealed in their minimum, maximum, median, lower and upper interquartile ranges. There is variability at each of the observation points in all the variables as shown in the interquartile range greater than zero. Consequently, the parameter estimates reveal a variation over space for all the independent variables. Specifically the coefficient of the variable size (m^2) reveals that a property within a particular precinct of the study area commands a price of R-912.25 (m^2), while a property in another area can command a price of R21276.2 (m^2).

Table 5.8 GWR model coefficients

Variable	Minimum	Lwr quartile	Median	Upr quartile	Maximum
Constant	-15908583.8	-4862899.9	-2960455.68	-1098811.44	3423422.6
Beds	-385015.12	92818.989	201558.93	348828.24	1330113.9
Quality	-145561.81	390412.31	604763.86	924764.78	2110049.7
Condition	-979714.88	-86401.19	159908.49	288774.81	947392.73
Storey	-249036.39	578108.56	919349.10	1209415.02	2648727.0
Bld_style	-3286973.38	-395034.91	-130595.48	7251.379	749385.57
View	-57925.422	175253.51	277357.00	514740.72	2368897.9
RMOS	-33738.455	-1588.139	6884.8675	18281.17	57783.97
Size	-912.24872	9246.441	11502.750	13517.49	21276.19
Pool	-40809.308	4625.066	6875.634	12227.99	28537.57

The negative sign on the lower end of the property price is somewhat counterintuitive. This kind of scenario was observed in the study of McCluskey *et al*. (2013: 257) and might be the effect of a larger house which might require modernisation. Details about the state of the property can only be verified from the property agents and tax assessors. The building style (Bld_style) has similar results which range from R-3286973.4 to R749385.6 which is linked to an increase in value of a property with a good building style in one area while property in another is low with a negative sign. Again, the parameter estimates for storey shows that "*ceteris paribus*" it sold from a range of as little as R-249036.39 at one location and R2648727.0 more at another location of Cape Town. The negative estimates suggest that the value of a storey building in Cape Town is highly dependent on the location. The result shows the benefit of using a non-stationary model over the stationary–coefficient model as it reveals prices of properties from different neighbourhoods/locations within a city.

The analysis of the autoregressive models (SEM and SLM) maximum likelihood estimates are presented in Table 5.9 and Table 5.10. The property size and swimming pool variables are significant in both the SEM and SLM. These variables are found to be significant in the linear, semi-log and log-log models which reveal their relative contribution to property prices in this study. Again the variables storey_2 and storey_3 are significant in both SEM and SLM.

Table 5.9 SEM model coefficients

Variable	Coefficient	Std.Error	z-value	Probability
Constant	1.53E+06	335846	4.5419	0.00001
Beds1	-291551	329299	-0.885369	0.37596
Beds2	-143666	105545	-1.36118	0.17346
Beds4	158117	67215.9	2.35238	0.01865
Beds5	470711	103770	4.53612	0.00001
Beds6	613813	172085	3.56693	0.00036
Beds7	622545	366990	1.69635	0.08982
Beds8	4.27E+06	592032	7.20849	0.00000
Beds9	49344.7	1.10E+06	0.0446578	0.96438
Beds10	2.60E+06	1.54E+06	1.69086	0.09086
Q_Poor	-339016	945311	-0.358629	0.71987
Q_Fair	88439.3	328816	0.268963	0.78796
Q_Good	439565	88255.2	4.98061	0.00000
Q_Vgood	2.23E+06	286146	7.78858	0.00000
Q_Exec	1.71E+06	267641	6.40342	0.00000
C_Poor	-410037	798722	-0.513367	0.60769
C_Fair	-233502	261295	-0.893635	0.37152
C_Good	342850	87533.5	3.91678	0.00009
C_Excell	206789	157026	1.31691	0.18787
Storey_2	703636	66283.8	10.6155	0.00000
Storey_3	1.74E+06	189648	9.19586	0.00000
S_Subecon	23409.4	901949	0.0259543	0.97929
S_Unconven	1.17E+06	189181	6.18139	0.00000
S_Geor/victor	208.885	252356	0.000827739	0.99934
S_Cape dutch	50962.1	434239	0.117359	0.90658
S_Maisonette	-251402	264558	-0.950274	0.34197
S_Med/t	-437594	743901	-0.588241	0.55637

V_ Partially obs.	171900	141046	1.21876	0.22294
V_Below aver	17719.5	361934	0.0489579	0.96095
V_Above aver	388837	75249.3	5.16732	0.00000
V_Panoramic	1.06E+06	111711	9.46288	0.00000
V_Excell	1.54E+06	232541	6.61184	0.00000
SA1	-40819.4	84479.5	-0.483187	0.62896
SA2	-71762.8	82070.9	-0.874401	0.3819
SA3	-54129.9	83188.3	-0.650691	0.51525
SA5	226758	85727.6	2.64509	0.00817
Size	9628.1	475.11	20.265	0.00000
Pool	6497.31	1727.26	3.76162	0.00017
Lambda	0.920082	0.00833279	110.417	0.00000

They are also significant in the linear, semi-log and log-log models with appropriate signs. The results also shows that properties with more than three bedrooms contribute more to property prices than properties with less than three bedrooms in Cape Town. In all there is high degree of similarities in the number of significant variables in SEM and SLM including the hedonic regression models. The negative values in some of the variables including bedroom, quality and condition of properties, building style and semi-annual time trend is contrary to the *a priori* expectation, but might be due to the law of diminishing marginal utility, i.e. an increase to the number of bedrooms from one to two should ordinarily be an additive to property price but at a depreciating rate. If combined with other variables, although within limits, multicollinearity might influence the results and provide incorrect absolute values which can only be assessed at individual attribute level. This would normally occur where insufficient data points exist that represent a particular variable.

Consequently, while properties that have poor (q_poor) and fair (q_fair) quality should ordinarily attract less additive and significant contribution to property prices, the same could not be said of properties in excellent condition (c_excell) in SLM which should primarily be additive towards value but at a depreciating rate. Importantly, the spatial lag term in Table 5.10 is positive and significant indicating that property prices in Cape Town are strongly influenced by the prices of nearby properties in line with Tobler's (1979) first law of geography to wit nearby properties are more related than properties that are far apart.

Table 5.10 SLM model coefficients

Variable	Coefficient	Std.Error	z-value	Probability
W_Assessed_v	0.827761	0.00979132	84.5403	0.00000
Constant	-1.76E+06	110251	-15.9613	0.00000
Beds1	-228659	339720	-0.673079	0.5009
Beds2	-33619.3	108491	-0.30988	0.75665
Beds4	144735	69112.7	2.09419	0.03624
Beds5	443680	107017	4.1459	0.00003
Beds6	718752	176559	4.07089	0.00005
Beds7	743840	380182	1.95654	0.0504
Beds8	4.26E+06	611499	6.96035	0.00000
Beds9	68096.1	1.14E+06	0.0596574	0.95243
Beds10	1.98E+06	1.62E+06	1.22786	0.2195
Q_Poor	-717308	988278	-0.725816	0.46795
Q_Fair	-123578	341419	-0.361954	0.71739
Q_Good	317614	89845.9	3.5351	0.00041
Q_Vgood	2.20E+06	297192	7.39704	0.00000
Q_Exec	1.80E+06	274888	6.53343	0.00000
C_Poor	242623	835536	0.29038	0.77153
C_Fair	-127721	270466	-0.472226	0.63677
C_Good	205997	89181	2.30987	0.0209
C_Excell	-21753.2	154038	-0.14122	0.8877
Storey_2	645417	65391.9	9.87	0.00000
Storey_3	1.96E+06	191746	10.2213	0.00000
S_Subecon	11321.5	933576	0.012127	0.99032
S_Unconven	1.40E+06	189226	7.38238	0.00000
S_Geor/victor	403296	255638	1.57761	0.11466
S_Cape dutch	-45335	449063	-0.100955	0.91959
S_Maisonette	46299.9	269814	0.1716	0.86375
S_Med/t	-359214	720496	-0.498566	0.61809
V_ Partially obs.	15109.4	138914	0.108768	0.91339
V_Below aver	35794.4	371943	0.0962362	0.92333
V_Above aver	285231	66366.5	4.29782	0.00002
V_Panoramic	995343	91021.5	10.9352	0.00000
V_Excell	1.43E+06	228282	6.28554	0.00000
SA1	-30684.4	87363.6	-0.351226	0.72542
SA2	-69801.9	85062.3	-0.820598	0.41188
SA3	-47627.4	86169.1	-0.552721	0.58045
SA5	200272	88017.6	2.27536	0.02288
Size	7917.64	461.78	17.1459	0.00000
Pool	6750.79	1720.98	3.92265	0.00009

5.4 Prediction of property prices with support vector machines, M5P trees and additive nonparametric regression models

This section estimates property prices with other non- or semi-parametric regression models. The analysis begins with the more recently introduced SVMs and M5P trees into the field of mass appraisal of properties which are been extended using the Cape Town property data. This is closely followed with the additive nonparametric regression that utilises an iterative backfitting algorithm which reduces multivariate regression to successive simple bivariate regressions (Bin, 2004: 70). The results also show the PLK model to perform poorly while the RBF marginally outperform the PLK model in all dataset (2055526), training (2085642) and testing (1994365) datasets. The MAE also reveals the NPK, SVMs to be closer in terms of price prediction to the actual assessed values in all data and the stratified datasets. Specifically, the MAE reveals an error of 905978 for all data, 857370 for training dataset and 740055 for the test dataset.

Table 5.11 summarised the results of the SVMs with three kernel functions. The variability in prices of properties has been effectively explained by the three kernel functions as revealed by the goodness of fit. Specifically, in the all dataset about 57.8%, 67.9% and 60.9% of the variation in property prices was explained by the polynomial kernel (PLK), normalised polynomial (NPK) and radial basis (RBF) kernels. The RMSE accuracy test shows the NPK support vector machines to perform optimal in terms of prediction and accuracy in the all dataset (1820201), training (1782006) and testing (1582401) datasets in comparison to the other models. The results also show the PLK model to perform poorly while the RBF marginally outperform the PLK model in the all dataset (2055526), training (2085642) and testing (1994365) datasets. The MAE also reveals the NPK, SVMs to be closer in terms of price prediction to the actual assessed values in the whole and the stratified datasets. Specifically, the MAE reveals an error of 905978 for the all data, 857370 for training dataset and 740055 for the test dataset.

Table 5.11 Goodness of fit and performance comparison of PLK, NPK and RBF kernels

Performance measures		R^2	Mean ratio	Median ratio	PRD	COD	MAE	RMSE
SVM all data	PLK	0.5784	1.0219	0.9989	1.1206	23.58	1143399	2114430
	NPK	0.6799	1.0138	1.0018	1.0866	18.05	905978	1820201
	RBF	0.6098	1.0182	1.0019	1.1165	21.63	1074495	2055526
SVM train data	PLK	0.5857	1.0275	0.9995	1.1194	23.75	1139552	2136034
	NPK	0.7061	1.0169	1.0013	1.0801	17.24	857370	1782006
	RBF	0.6142	1.0242	1.0001	1.1174	22.02	1075499	2085642
SVM test data	PLK	0.5909	1.0058	0.9985	1.1180	22.24	1112048	1995947
	NPK	0.7261	1.0110	1.0019	1.0691	14.41	740055	1582401
	RBF	0.6073	1.0056	1.0018	1.1258	21.47	1088171	1994365

Other test statistics in line with IAAO guidelines for mass appraisal reveal the NPK, support vector machines to achieve a COD of 18.05%, 17.24% and 14.41% for the whole data, training and testing datasets, respectively. The test dataset of the NPK performed best in terms of uniformity and horizontal dispersion accuracy across the total sales. Similar trends follow the other accuracy statistics including the median ratio, mean ratio and the PRD. The result shows the NPK support vector machines as the best performing model compared to the other kernel functions in terms of lower error in prediction, vertical equity and uniformity.

The study also evaluated the performance of other non- and semi-parametric models. The results in Table 5.12 summarised the performance of M5P trees and ANR. The variables (except the dummy variables) were log transformed in this analysis. Though the structure of the two models is dissimilar, particular attention was paid to model performance in line with their COD and PRD. The results reveal the ANR to outperform the M5P in terms of the COD and PRD. Results also show that in terms of the MAE, the ANR predicts prices that are closer to the assessed values and it also produced lower assessment error as revealed in their RMSE. The ANR model is similar to the baseline (semi-log regression) model but performs better than the linear and log-log models.

Table 5.12 Goodness of fit and performance comparison of M5P trees and ANR modelling

Performance measures	M5P trees			ANR		
	All data	Training	Testing	All data	Modelling	Testing
R^2	0.5589	0.5682	0.5884	0.693	0.6200	0.6154
Median ratio	1.0647	1.0537	1.0744	1.028	1.0291	1.0268
Mean ratio	1.0929	1.0965	1.0896	1.044	1.0364	1.0477
PRD	1.0929	1.0965	1.0896	1.1022	1.1018	1.1024
COD	26.59%	27.62%	26.02%	22.11%	22.28%	21.74%
MAE	1259703	1257336	1250357	1092182	1081378	1096814
RMSE	2070205	2096862	1884810	1951989	1854770	1992225

5.5 The influence of ANNs training algorithms in mass appraisal

The structure of the ANNs is different from the regression based techniques. There is no *a priori* pre-specification of the relationship between the dependent and independent variables. The variables used for ANNs training and testing are identical but having differing composition as previously noted. Four ANNs training algorithms including BP, SCG, PBCG and LM were used in this assessment. The results are presented in Table 5.13. The performance of models reveals a very good square correlation between the actual assessed values and the estimated values for the all data, training and testing datasets. For the BP training algorithm, the R^2 reveals that the ANNs architecture explains 67.9% of prices for all data, 66.2% for training data and 77.5% for testing data. There is improvement in the performance of the SCG, PBCG and LM relative to the BP in explaining the variation in property prices. Specifically, the SCG algorithm explains 75.9%, 75.6% and 76.5% of variation in property prices for all data, training and testing data. Also the R^2 statistic measures for PBCG revealed 75%, 75.1% and 74.9% variation in property prices for all data, training and testing data. The LM revealed variance in property price as shown in their R^2, 75.8%, 79.7% and 67.2% for all data, training and testing datasets, respectively.

The MAE accuracy test statistic revealed the LM algorithm to predict property prices closer to the assessed values than all other training algorithms for all, training and testing data (239298, 243832 and 228724). The BP training algorithm performed worse in the all data and testing data (1057708 and 1013821) but have slight improvement in the training data (1288025) as revealed in the MAE accuracy test. BP algorithm slightly outperform the SCG

Table 5.13 Goodness-of-fit measurements and prediction comparison of BP, LM, PBCG and SCG algorithms

Performance measures		R^2	Mean ratio	Median ratio	PRD	COD	MAE	RMSE
SCG	All data	0.7590	1.0263	0.9941	1.0275	11.77	256469.9	397494.6
	Train data	0.7558	1.0251	0.9940	1.0277	11.82	1762084	2889345
	Test data	0.7653	1.0291	0.9965	1.0272	11.62	253126.3	407919.9
PBCG	All data	0.7504	1.0276	0.9978	1.0257	11.75	255564.8	527264.9
	Train data	0.7509	1.0279	0.9979	1.0262	11.85	257894.6	571605.3
	Test data	0.7495	1.0269	0.9978	1.0245	11.50	250131.6	405444.6
LM	All data	0.7583	1.0129	0.9957	1.0229	10.77	239297.9	513824.2
	Train data	0.7973	1.0139	0.9936	1.0239	10.98	243832.4	559611.0
	Test data	0.6716	1.0106	1.0003	1.0207	10.26	228723.7	386524.8
BP	All data	0.6789	1.0461	1.0278	1.0829	25.14	1057708	1773281
	Train data	0.6618	0.8262	0.7951	1.0180	28.76	1288025	2042443
	Test data	0.7749	1.0859	1.0413	1.0606	24.16	1013821	1410072

in the training data. The RMSE accuracy test shows the BP algorithm to perform poorly in all data and test data in comparison to the other training algorithms. The BP algorithm also marginally outperforms the SCG in the training dataset. The LM reveals a lower RMSE in the training and testing dataset, thus showing that in terms of model prediction and accuracy it is better than other training models.

The results of the different ANN training algorithms when viewed in the light of the IAAO approved guidelines for mass appraisal reveal that apart from the BP training algorithm, all other models demonstrate consistency between lower valued and higher valued properties. The PRD for all other models and the training data of BP reveals that it lies between 1.01–1.03, depicting that neither the lower nor the higher valued properties are favoured. However, the result shows that the BP algorithm performs better in one scenario – the training data than the PRD of training data of other models. Again, analysis shows that the LM algorithm performs better in the training and testing datasets in terms of their COD statistic but the BP algorithm performs poorly in terms of uniformity and horizontal or random dispersion (COD). The analysis reveals a high similarity of results among the different algorithms, therefore to arrive at a definite conclusion on which training algorithm perform optimal a reliability ranking order as used in McCluskey, *et al*. (2013: 258) was employed. The analysis was carried out on the R^2, PRD, COD, MAE and RMSE for all data, training and testing datasets (Table 5.14).

Table 5.14 Performance of algorithms and reliability ranking order

Training algorithms	R^2	PRD	COD	MAE	RMSE	Overall rank
	All data					
SCG	1	3	2	3	1	2.0
PBCG	3	2	1	2	3	2.2
LM	2	1	2	1	2	1.6
BP	4	4	4	4	4	4.0
	Training data					
SCG	2	4	2	4	4	3.2
PBCG	3	3	3	2	2	2.6
LM	1	2	1	1	1	1.2
BP	4	1	4	3	3	3.0
	Testing data					
SCG	2	3	3	3	3	2.8
PBCG	3	2	2	2	2	2.2
LM	4	1	1	1	1	1.6
BP	1	4	4	4	4	3.4

The results of reliability ranking order in Table 5.14 reveal the LM training to outperform all other models in this study. The LM algorithm has the properties of gradient descent and Newton speed which enables it to train the ANNs faster than the most widely used BP algorithm. Although the SCG and PBCG algorithms perform well in this study they nonetheless could not surpass the LM training algorithm. The next section concerns the use of meta-heuristic algorithm in optimising and training the neural networks in mass appraisal of properties.

5.6 Combining GABP and PSOBP in weight optimisation and ANNs training

The advance in technology has led to a number of improvement to the performance of ANNs. The preceding section revealed the importance of other training algorithms in the prediction of property prices. There has been a gradual shift from the use of standalone algorithms to combining two or more algorithms. This section presents the result of the analysis involving the combination of GA and BP on the one hand and PSO and BP on the other. It should be understood that the platform used in building the hybrid systems is constrained to fewer measures namely, the R^2, MAE and RMSE. The lack of detailed results inhibited a robust analysis relative to other models used in this study. This notwithstanding, since the R^2, MAE and RMSE accuracy test statistics have proved useful in measuring the performance of models in many studies, these were utilised to compare performance of the two hybrid systems. The results of the hybrid models are compared to those of standalone artificial neural networks relative to the available performance measures to assess their levels of prediction accuracy.

5.6.1 Generated attributes weights by PSO, GA and BP algorithms

The study of McCluskey & Anand (1999: 226) reports that attribute weights reveal the relative importance of individual attributes in a model. The higher the attribute weight of a given sample the more relevant the attribute is considered, conversely the lower the weight the less important it is and less consideration is given to the attribute. The PSO and GA are used to generate an attribute's weight (fittest solution) for the problem. Furthermore, the fittest solution will results in the lowest mean absolute error and root mean squared error on the whole (all) data, training and testing datasets. Table 5.15 reveal the results of the discovered weights by PSO and GA with BP algorithms on the whole dataset.

Table 5.15 Revealed attribute weights by PSOBP and GABP

Attribute	GABP	PSOBP	Attribute	GABP	PSOBP
Beds	0.6325	0.6234	Submkt53	0.2248	0.2405
Quality	0.8880	0.8634	Submkt55	0.3958	0.4011
Condition	0.5699	0.5646	Submkt56	0.1697	0.1888
Storey	0.0419	0.0688	Submkt64	0.0100	0.0294
Bld_style	0.2035	0.2205	Submkt65	0.1203	0.1423
View	0.4848	0.4847	Submkt66	0.0511	0.0774
RMOS	0.1630	0.1825	Submkt67	0.2511	0.2652
Size	1.0000	0.9685	Submkt68	0.2671	0.2802
Pool	0.5672	0.5621	Submkt69	0.7835	0.7652
Submkt48	0.9661	0.9366	Submkt70	0.3937	0.3991
Submkt50	0.0343	0.0616	Submkt73	0.5496	0.5456
Submkt52	0.7062	0.6926			

The results reveal that the GABP found property size (1.00) and submarket 48 (0.966) to be the most significant attributes in this sample, while number of storeys (0.042) and submarket 64 (0.01) were assigned lesser weights depicting that they are less significant. The PSOBP also determined the property size (0.969) and submarket 48 (0.937) to be the most significant attributes while storeys (0.069) and submarket 64 (0.029) are the least significant attributes. For the purpose of comparison, a benchmark of 20% was used to assess the relative importance of property attributes in this study. The result reveals the PSOBP to have a marginally larger number of important variables (17) than the GABP (16). Furthermore, the weights determined by GABP and PSOBP are very similar with the two most important attributes also found to be significant in the linear, semi-log and log-log models. This interpretation does not construe a similarity in attribute weightings of the hybrid systems to the coefficients in the regression models. The output weights generated through the internal workings of the hybrid models having the black box nature of the neural networks cannot be taken as price adjustment but it provides a glimpse into the relative importance and contribution of attributes.

Table 5.16 Goodness of fit and performance comparison of GABP and PSOBP modelling

Performance measures	GABP			PSOBP		
	All data	Training	Testing	All data	Training	Testing
R^2	0.592	0.597	0.601	0.595	0.589	0.601
MAE	1194758	1216373	1234376	1201939	1213889	1234376
RMSE	1989284	2035167	1859916	1984238	2046221	1859916

The results in Table 5.16 reveal the performance of the hybrid systems. The performance of the GA depicted by the R^2 shows that the architecture explains 59.2% (all dataset), 59.7% (training data) and 60% (testing dataset) of the variability in property prices while the PSO explains 59.5% (whole dataset), 58.9% (training data) and 51.5% (testing data) of the property prices. The RMSE accuracy test reveals the PSOBP to achieve 1984238 in the all dataset, 2046221 in the training set and 1859916 in the test data. The PSOBP achieve a RMSE of 1989284, 2035167 and 1859916 for all data, training and testing data respectively. The MAE value reveals the GABP to predict values that are closer to the actual assessed values than the PSOBP. In all the results of GABP and PSOBP built in this study are similar. However when the results of the hybrid systems are compared to the standalone BP, LM, PBCG and SCG, the hybrid systems generated higher errors. This is, however, contrary to the *a priori* expectation and requires further examination. It is only in the training set of SGP and BP that the hybrid systems outperformed. The expectation is that the hybrid systems should provide a higher level of transparency than the standalone algorithm but this was not achieved in this study.

5.7 Effective comparison of the performance of models

Previous studies are not conclusive in terms of model superiority. Pertinently, the observations raised in McCluskey *et al*. (2012: 285) on the need to ask the right questions when comparing models to wit “which modelling approach meets the rigorous standards of transparency, stability of output, predictive ability and defensibility for the mass appraisal industry?” is worthy of consideration in the selection of the best model for the Cape Town property market. The preceding analyses reveal the following models to outperform those in similar category. The baseline model is the semi-log; the best performing model built to tackle spatial effects in this study is the spatial error model; and, thirdly the overall best support vector kernel is the normalised polynomial kernel optimise with sequential minimal optimisation. The optimal artificial neural networks training algorithm is the Levenberg-Marquardt and finally the additive nonparametric regression.

In all, five models are chosen from which the best model for mass appraisal of properties in the Cape Town property market is selected. The hybrid system is not included in this analysis because the standalone artificial neural networks relatively outperform the GABP which marginally outperform the PSOBP. The tests used for comparison are the model performance and reliability ranking order, model explainability ranking order and prediction accuracy of a model within 10 and 20% of the actual assessed values (McCluskey *et al*., 2013: 259). The model that performs best in all measures is selected as the Cape Town appraisal model. It should be noted that most models do not meet the IAAO guidelines in terms of the accuracy of the model outputs, but rather serve as a baseline of comparison for the methods described in this study (Borst, 2006: 13). This might be the consequences of model misspecification and outliers.

The first test is to assess the level of prediction accuracies which falls within 10 and 20% of the assessed values as used by Thibodeau (2003: 17) and McCluskey, *et al*. (2013: 255). The results are presented in Table 5.17. The overall results reveal that semi-log and ANR models fail the minimum standard of 50% prediction accuracy within 10% of the assessed values. Only the SEM, NPKSVMs and LMANNs marginally achieve the benchmark of 50%. Nevertheless since the analysis is meant to find a model that perform best relative to others; these models have a good predictive performance. The ANR performed worst with 39.4% of the predicted values falling within 10% of the assessed values. There is a slight improvement in semi-log and ANR when the benchmark was increased to prediction within 20% of the assessed values. In all, the LMANNs performed best in this test. The model has a good prediction accuracy of 53.1% property appraised within 10% and 57.1% predicted values falling within 20% of the assessed values.

Table 5.17 Model prediction within 10 and 20% of actual assessed values

Model	Percent within 10%	Percent within 20%
Semi-log	43.4%	50.5%
SEM	50.6%	52.1%
NPK-SVMs	50.7%	55.1%
ANR	39.4%	49.6%
LM-ANNs	51.1%	57.1%

The second test employed is the performance and reliability ranking. The R^2, median and mean ratios, COD, PRD, MAE and RMSE accuracy measures were used. The accuracy measures were selected because they are found in all platforms used in the analysis of the selected models. The results are summarised in Table 5.18.

Table 5.18 General model performance and reliability ranking order

Accuracy measures	Semi-log Rank	SEM Rank	NPKSVMs Rank	ANR Rank	LMANNs Rank
R^2	3	2	4	5	1
Median ratio	4	2	3	5	1
Mean ratio	4	2	3	4	1
PRD	4	2	3	4	1
COD	4	2	3	4	1
MAE	4	3	2	5	1
RMSE	4	2	3	5	1
Overall rank	3.86	2.14	3.00	4.57	1.00

The results presented in Table 5.18 are tied to the Cape Town property data. The comparative analysis reveals the Levenberg-Marquardt trained artificial neural networks (LMANNs) to outperform all other models. The LMANNs has both gradient descent search method and Newton speed which enhances its performance. This is closely followed by the SEM. The SEM model has the history of good performance which is made possible by the specification of spatial dependence of properties in the model. The NPKSVMs ranked third while the semi-log model ranked fourth. The overall result favours the LMANNs as the best in terms of predictive accuracy but it should be noted that within the mass appraisal field, equally important is the ability to support the assessed/estimated values before a tribunal (McCluskey *et al*., 2012: 284). Although such challenges against the model rarely occur, the detail estimates predicted should adequately be comprehensible to the appraiser to such an extent that in any challenge, it could be explained explicitly.

Table 5.19 provides the results of the explicit ability of a model to explain details which is very important in mass appraisal setting. The hedonic regression based model has the history of being used within the mass appraisal environment despite its shortcomings, because of the simple and consistent manner in which it provides statistical evidence that can help in objective judgement about quality of the predictive model in terms of R^2, adjusted R^2, *F*-statistics, *t*–value and levels of significance of variables. Again McCluskey, *et al*.

(2013: 258) noted that the traditional hedonic regression models are preferred because of the single set of parameter estimates which is consistently applied. These qualities of the HRMs give it an edge over other models used in this comparative analysis, hence, the semi-log

Table 5.19 Model explicit explainability ranking order

Measurement Criteria	Semi-log Rank	SEM Rank	NPKSVMs Rank	ANR Rank	LMANNs Rank
Simple	1	2	4	3	5
Consistent	1	2	4	3	5
Transparency	2	1	4	3	5
Locational	2	1	4	3	5
Applicability	2	1	4	3	5
Mean ranking	1.60	1.40	4.00	3.00	5.00

model occupy unique positions one and two as shown in Table 5.19. The SEM is similar to the traditional hedonic regression based in terms of simplicity and consistency but has a marginal increase in terms of transparency because of the way it vary its maximum likelihood estimates. The ANR is equally similar to the traditional regression based models but marginally performed below the semi-log model. The ANNs and SVMs do not provide such details and are less transparent than the other techniques.

In terms of explicit nature of location, the SEM include an absolute measure (x, y coordinates) of location and also permits parameter estimates to vary based on location of properties. The traditional models (semi-log) is not inherently locational (McCluskey *et al.*, 2013: 259), the same with the ANR but locational dummies are used to reflect their influence. The ANNs and SVMs have abilities to recognise patterns; hence, locational dummies are equally used. The most important measure is the applicability of models in the mass appraisal environment. It is not enough for a model to accurately predict values, but also to what extent can the model be used within the mass appraisal environment. Table 5.18 revealed the LMANNs to be the best performing model in terms of prediction accuracy but results in Table 5.19 negates this finding due to the non-transparent nature of the estimates predicted by the model. The black box nature of the ANNs and SVMs makes it difficult to implement within the mass appraisal environment. The other three models (ANR, semi-log and SEM) can be implemented for mass appraisal but in the context of this research the mean ranking favours the SEM.

5.8 Chapter summary and conclusion

This chapter provided a comprehensive analysis and discussion of results on the prediction accuracy of all models. Several evaluation criteria were used to assess the level of their performance. The R^2 used only reveal the explainability of variance in property prices of the models but other accuracy test statistics such as MAE, RMSE and those acceptable to the IAAO including the PRD and COD revealed much detail of the performance of models. The results clearly demonstrate the practicability of building a hybrid model of PSO with BP to optimise and train ANNs. The results also show the superiority of training the ANNs with other algorithms namely LM, PBCG and SCG. The chapter also show the efficacious use of logarithmic transformation of property variables to enhance model performance and applied spatial varying and weighted models in the assessment of property prices in the South African property market context. In summary, the LMANN has been found to outperform all other models used in terms of prediction accuracy but when considered in relation to simplicity, transparency, consistency and applicability within the mass appraisal environment, particularly, for property tax administration the model falls short.

Within the mass appraisal environment a model is adjudge worthy of use if it is simple, consistent and transparently provide details that can aid appraisers defence before a tribunal. These requirements are virtually absent in the ANNs and SVMs but the ANR, semi-log and SEM provides these details with the SEM surpassing them because of its ability to handle spatial dependence. Therefore the SEM is favoured for mass appraisal of properties in the Cape Town property market. ANNs and SVMs, however, still has a good history of ease of use, cost effectiveness and capability of handling nonlinear data which placed them at an advantaged position of being used as a check on the estimates produced by the regression based models. An area of concern in this study is the relatively high PRD and COD results which might be caused by misspecification and outliers. It should be noted, however, that the aim of this study is not to ensure an absolute accurate model, but rather a comparison of different models against each other.

6. CHAPTER SIX

6.0 Conclusion and Recommendations for Future research

6.1 Introduction

Mass appraisal of properties is gradually been accepted in most developing countries. The traditional hedonic model is the most widely used technique used to advise mortgage institutions, property tax authorities and portfolio managers. The purpose of this study is to build an improved technique for modelling property prices in Cape Town. Several techniques of modelling property prices exist, but the concern is that there has not been a definite consensus on which technique could be effectively used in all jurisdictions and within geographical context. This is partly caused by variability in property data and quality necessitating the need to develop a suitable technique with the Cape Town property data. Consequently, the research had the following objectives as stated earlier:

- ✓ Establish a baseline regression model for the Cape Town property market;
- ✓ Assess the performance of flexible regression models in mass appraisal of properties;
- ✓ Investigate the influence of ANNs training algorithms in mass appraisal of properties;
- ✓ Build a hybrid model from existing artificial neural networks (ANNs) to create a more effective algorithm for mass appraisal/valuation of properties; and
- ✓ Compare the performance of models in terms of predictive ability, explainability, defensibility and use within the mass appraisal environment.

The chapter is divided as follows: Section 6.1 contained an introductory review of the findings, Section 6.2 will address each objective to evaluate the extent to which it has been achieved, Section 6.3 provides some concluding remarks on the research and findings, Section 6.4 indicate the practical implications of the research and Section 6.5 makes some recommendations for future research.

6.2 Realisation of study objectives

6.2.1 Establish a baseline regression model for the Cape Town property market

A model that forms the basis of comparison throughout the study was established. The different functional specification of the traditional hedonic regression was used in achieving this objective. A RESET test advocated by Peterson & Flanagan (2009) and adopted by McCluskey, *et al.* (2012) to justify the used of linear, semi-log and log-log models was undertaken. The three models having different structures were tested against the industry and standard benchmarks to reveal the semi-log regression models as the baseline model for the Cape Town property market.

6.2.2 Assess the performance of flexible regression models in mass appraisal of properties

A considerable body of knowledge on several flexible regression models exists. These models are sometimes referred to as non- or semi-parametric regression models. The models are GWR, SEM, SLM, SVMs, M5P trees and ANR. The GWR was used to account for spatial heterogeneity and local variation in property prices. The SEM and SLM were used to account for spatial dependence. The analysis revealed the SEM as the most preferred of all the spatially weighted regression models in this assessment.

The SVMs use differing kernel functions including the polynomial, normalised and radial basis function kernels. The sequential minimal optimisation was used to optimise and supply the support vectors. The data for all models in this category was stratified into training/modelling and testing sets. In carrying the analysis a 100% data signifying all (whole), a 70% data for training/modelling and a 30% data for testing was used. Using standard and industry test statistics the result favours the normalised kernel support vector machines. The other analysis involving the M5P trees and ANR produced a result that favours the ANR. Consequently, the three models that performed well namely spatial error model, normalised polynomial kernel and the additive nonparametric regression were added to the number of models used for general comparative analysis. These models were previously unreported in any known South African property market analysis.

6.2.3 Investigate the influence of ANNs training algorithms in mass appraisal of properties

The study introduces other unreported artificial neural networks training algorithm in mass appraisal of properties. These are PBCG and SCG algorithms. They were tested on the Cape Town property data along with the BP and LM training algorithms. Using the industry and standard benchmark accuracy statistical tests, the results reveal that PBCG, LM and SCG algorithms outperform the traditional BP algorithm in this study with the overall performance been credited to the LM training algorithm. This confirms the research of El Hamzaoui & Perez (2011) and elaborates on the findings in terms of the detailed analysis and comparison to other available methods.

6.2.4 Build a hybrid model from existing artificial neural networks (ANNs) to create a more effective algorithm for mass appraisal/valuation of properties

A hybrid systems of GABP and PSOBP designed to optimise and train the ANNs was developed in this study. The hybrid system was used to generate attribute weights from global and local spaces and train the ANNs. The result shows that in terms of predictive accuracy, the test statistics reveal similar results between the hybrid systems with the GABP marginally performing better than the PSOBP. When the results were compared to the standalone ANNs, the standalone training algorithms performed better. This was contrary to the *a priori* expectation because of the belief that two or more algorithms should produce better results and enhance transparency, but this was not achieved in this study. The limitation of the platform in producing more results inhibited a robust assessment in terms of the performance of the hybrid systems relative to the standalones in terms of their COD's and PRD's.

6.2.5 Compare performance of the models in terms of predictive ability, explainability, defensibility and used within the mass appraisal environment

This objective brings together selected models that performed best in their respective category for comparison. The models are the baseline semi-log model, the spatial error model, the normalised polynomial kernel, the additive nonparametric regression and the Levenberg-Marquardt artificial neural networks. They were tested using the prediction accuracy within 10 and 20% of the assessed values, the performance and reliability ranking order and the explicit and reliability ranking order. While the LMANNs perform best in the first two tests, the model was rejected on the basis of non-transparency and applicability within the mass appraisal environment. The same thing follows the NPKSVMs which trailed very closely behind the LMANNs in the first test but behind the SEM in the second test in terms of prediction accuracy. The results favour the SEM which though in the context of predictive accuracy was below the LMANNs but preferred as the best in the contexts of simplicity, consistency, locational, transparency and applicability in the mass appraisal domain.

6.3 Conclusion

The need to find a model that predicts values of properties devoid of the many parametric restrictions of the hedonic regression model has been a long standing issue within the mass appraisal industry. The model has traditionally been utilised for the appraisal of properties but has over the years been associated with a number of limitations including functionality and nonlinearity. Several non- and semi-parametric regression models including the ANNs, GWR, SAR, ANR, SVMs amongst others have been introduced to deal with the obvious weaknesses of hedonic regression. However, previous studies undertaken to compare performance of different modelling approaches with a view to ascertaining the superiority have not yielded the needed results. Primarily the major concerns are related to the data quality and the geographical contexts where the study is been undertaken.

The non- and semi-parametric regression models offer good modelling methodology that is not affected by the many limitations of the hedonic regression. The ANNs and SVMs have powerful pattern recognition properties that can efficaciously recognise complex value effects in property price analysis. The spatial models (GWR, SEM and SLM) have abilities to tackle spatial heterogeneity and dependence thus predicting property prices devoid of inefficient and inconsistent parameter estimates. The M5P trees designed to deal with functionality limitations has the ability to generate model trees in which a linear regression is estimated for each leaf on the basis of the observations that get to it. The ANR introduces a smooth spline to deal with the parametric rigidity of the hedonic regression. The models were all tested on a property dataset from Cape Town, South Africa. In each category the overall best model was selected for a comparative analysis on a set of criteria including prediction accuracy, simplicity, consistency, locational, transparency and applicability within the mass appraisal environment.

The recent developments in the machine learning field have tended towards the use of meta-heuristics and other training algorithms in ANNs to increase its capability and predictive performance. Whilst it is true that ANNs is flexible in handling nonlinearities and developing price models without having to create binaries or linearise variables through transformation; again whilst it is equally true that ANNs have relatively overcome training limitations with the use of nature inspired (GA and PSO) and other improved algorithms, they do however suffer from non-transparency in interpretation of coefficients. Again the trial and error procedure of ANNs can create slightly inconsistent and unreliable results each time the technique is used. The SVMs also having the machine learning paradigm do suffer the same fate of non-transparency in interpreting the coefficient estimates. The findings clearly show SEM, semi-log and ANR models to provide transparency because of the explicit coefficients they provide. The coefficients are particularly useful in defence when appraisal estimates are challenged. Therefore, the SEM though in terms of predictive accuracy performed below the ANNs but was preferred because it passes the benchmark test of simplicity, consistency, locational, transparency and applicability. The ability of the SEM to use a discrete spatial reference (*x* and *y* coordinates) for each property, unlike the semi-log and ANR, provides a premise for effective assessment as increase weights are assigned to nearer and more similar properties while lesser weights are assign to distant properties.

Finally the results of this study reveal that ANNs is a viable tool in mass appraisal of properties but will only support the view that it is a substitute to the hedonic regression and its variants if the black box is opened.

6.4 Practical application of the study

The goal of this research was to find a model from arrays of models that perform well in all the benchmark tests so that the model could be implemented in the assessment of properties in the South African property market. The South African property market particularly the city valuation office, Cape Town is the primary beneficiary of the outcome of this research. The most widely used techniques for property tax assessment, mortgage underwritings, and portfolio management is the HRMs. This method in this study is still relevant given its simplicity of use, consistency and reliability of estimates, applicability within the mass appraisal environment and transparency. However, the SEM model has achieved these qualities to a greater extent and offers more promise to the mass appraisal industry because of its ability to tackle the problem of spatial dependence. While a number of studies have found ANNs and SVMs to be useful and powerful tools at handling nonlinear data which is also evidenced in this study, they suffer a huge setback because of non-transparency of their estimates. This research provides substantial evidence that the SEM is a superior model.

This research provides a thorough and complete comparison of various high level appraisal methods, which provide testing on a single database in order to compare all on an equal base. To the knowledge of the researcher this has not been conducted in the past. In addition to this, the research provides an introduction of these models for use in South Africa, where even linear regression analysis is used to a very limited extent. It thus offers a basis for the practical application of the various methods, including an understanding of their workings, which could be used in the local market due to the familiarity of industry players with the case study that was reviewed.

The SEM, as the method pointed out to be best suitable for application, can be applied in practice with readily available software packages designed to handle the application of SEM. This software ranges from the open source packages including *GeoDa*, Spatial Analysis in Macroecology (SAM), and the *R* software to those of commercial licence including Matlab and STATA, amongst others. Implementing the SEM is useful particularly to organisations that have previously utilised the regression models in SPSS, and Number Cruncher Statistical Systems (NCSS) for mass appraisal of properties.

6.5 Recommendations for Future Research

Following the preceding discussion above, future research can be conducted to possibly improve the methodologies. Possible way of such improvements could be to: (i) use a logarithmic form of the response variable in the spatial models; (ii) use software that facilitates analysis where detailed results could be used to achieve industry and more accuracy test statistics in the hybrid systems. This will make it possible to ascertain their level of prediction accuracy relative to standalone ANNs; (iii) revisit the Cape Town property data in order to see what causes the problem of higher COD and PRD in some of the techniques used, i.e. accuracy of data collected, model specification errors caused by uncaptured attribute information, or market inefficiencies; (iv) data from other jurisdictions within the South African property market would be useful to determine the robustness of the findings. It will also be of great interest to utilise property data from other regions within the sub-Saharan Africa to ascertain the robustness of the models; and (v) improve the degree of explainability of ANNs and SVMs through a research that promote the opening of the black box (to see what is happening inside) so that parameter estimates can be viewed explicitly. Until this is done these models would continue to be regarded as techniques that supports and complements econometric analysis and not as substitutes.

Again, to add voice to what the study of McCluskey, *et al*. (2012: 286) opined having considered the computing abilities of the regression techniques and the ANNs that a hybrid of the duo should be formed such that the results from ANNs could be used to smoothing the output generated by the HRMs. This is because despite the high performance of other training algorithms in this study, the results did not make any difference with the back propagation training algorithm in terms of transparency. Additionally, while the hybrid systems of regression techniques and the ANNs are herein advocated, a similar development should be extended to the SVMs with the regression based techniques so that a meaningful comparison of these high profile models will be undertaken to advise the assessment community of the suitability of the hybrid systems in mass appraisal.

7. References

Acciani, C., Fucilli, V & Sardaro, R. 2011. Data mining in real estate appraisal: a model tree and multivariate adaptive regression spline approach, *AESTIMUM* 58, Giugno, pp. 27–45.

Al-Saqer, S.M & Hassan, G.M. 2011. Artificial neural networks based red palm weevil (Rynchophorus Ferrugineous, Olivier) recognition system, *American Journal of Agricultural and Biological Sciences*, vol. 6, no. 3, pp. 356–364.

Amarappa, S. & Sathyanarayana, S.V. 2014. Data classification using Support vector Machine (SVM), a simplified approach, *International Journal of Electronics and Computer Science Engineering*, vol. 3, no. 4, pp. 435–445.

Ambarish, G.M & Saroj, K.L. 2016. Neural network pattern classification and weather dependent fuzzy logic model for irrigation control in WSN based precision agriculture, *Proceedia Computer Science*, vol. 78, pp. 499–506.

Andrei, N. 2007. Scaled conjugate gradient algorithms for unconstrained optimization, *Computational Optimization and Applications*, vol. 38, issue 3, pp. 401–416.

Andrei, N. 2011. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization, *Bulletin of the Malaysian Mathematical Sciences Society*, vol. 34, no. 2, pp. 319–330.

Anglin, P. & Gencay, R. 1996. Semi-parametric estimation of hedonic price functions, *Journal of Applied Econometrics*. Vol. 11, pp. 633–648.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publisher.

Anselin, L. 2003. Spatial Econometrics, in *A Companion to Theoretical Econometrics*, Ed. Baltagi, B.H., Blackwell Publishing Limited.

Anselin, L. 2005. *Exploring Spatial Data with GeoDaTM : A Workbook*. http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf

Bai, Y, Zhang, H & Hao, Y. 2009. The performance of the backpropagation with varying slope of the activation function, *Chaos, Solitons and Fractals*, vol. 40, pp. 69–77.

Battiti, R. 1992. First- and second-order methods for learning: between steepest descent and Newton's method, *Neural Computation*, vol. 4, pp. 141–166.

Beale, E.M.L. 1972. A derivative of conjugate gradients, in FA Lootsma ed. *Numerical Methods for Nonlinear Optimisation*, Academic Press, London, pp. 39–43.

Besner, C. 2002. A spatial autoregressive specification with a comparable sales weighting scheme, *Journal of Real Estate Research*, vol. 24. No, pp. 193–211.

Bertini, I., De Felice, M. & Pizzuti, S. 2010. Combining back-propagation and genetic algorithms to train neural networks for start-up time modelling in combined cycle power plants, Conference: *ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning*, Bruges (Belgium), 28-30 April 2010, pp. 165 - 170

Bin, O. 2004. A prediction comparison of housing sales prices by parametric versus semi-parametric regressions, *Journal of Housing Economics*, vol. 13, no. 3, pp. 68–84.

Bitter, C., Mulligan, G. F. & Dall'erba, S. 2007. Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method, *Journal of Geographical Systems*, vol. 9, pp. 7–27.

Borst, R.A. 1991 Artificial neural networks: the next modelling/calibration technology for the assessment community? *Property Tax Journal*, vol. 10, no. 1, pp. 69–94.

Borst, R.A. 1995 Artificial neural networks in mass appraisal, *Journal of Property Tax Assessment & Administration*, vol. 1, no. 2, pp. 5–15.

Borst, R.A. 2006. The comparable sales method as the basis for a property tax valuation system and its relationship and comparison to geostatistical valuation models, *paper in the international congress on advances in mass appraisal methods*, Delft university of technology,

Borst, R.A. 2007. Discovering and applying locational influence pattern in the mass valuation of domestic real property, PhD thesis submitted to the faculty of engineering, University of Ulster, UK.

Borst, R.A. 2008. Evaluation of the Fourier transformation for modelling time trends in a hedonic model, *Journal of Property Tax Assessment*, vol. 6, issue 4, pp. 33–40.

Borst, R.A. 2009. Time-varying parameters: obtaining time trends in a hedonic model without specifying their functional form, *Journal of Property Tax Assessment*, vol. 6, issue 4, pp. 29–36.

Borst, R.A. 2012. A space-time model for computer assisted mass appraisal, *XLI Incontro di Studio del Ce.S.E.T.*, pp. 535–545.

Borst, R.A. 2014. *Improving mass appraisal valuation models using spatio-temporal methods*, International Property Tax Institute, Toronto, Canada.

Boser, B.E., Guyon, I., & Vapnik, V.N. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop of Computational Learning Theory*, Pittsburgh, ACM, vol. 5, pp. 144–152.

Bourassa, S.C., Cantoni, E. & Hoesli, M. 2007. Spatial dependence, housing submarkets and house prediction, *Journal of Real Estate Finance & Economics*, Vol. 35, pp. 143–160.

Bourassa, S, Hamelink, F, Hoesli, M, MacGregor, B. 1999. Defining housing submarkets, *Journal of Housing Economics*, vol. 8, pp. 160 – 183.

Bourassa, S.C., Cantoni, E. & Hoesli, M. 2010. Predicting house prices with spatial dependence: a comparison of methods, Journal of Real Estate Research, Vol. 32, no. 2, pp. 143–160.

Bremermann, H.J. 1962. Optimisation through evolution and recombination. In Yovits, M.C., Jacobi, G.T. & Goldstine, G.D. (Editors), *Self-organisation Systems*, Spartan Books, pp. 93–106.

Brunsdon, C., Fotheringham, A.S. & Charlton, M.E. 1996. Geographically weighted regression: a method for exploring spatial nonstationarity, *Geographical Analysis*, vo. 28, no. 4, pp. 281–298.

Casetti, E. 1972. Generating models by the expansion method: applications to geographical research, *Geographical Analysis*, vol. 4, pp. 81–91.

Casetti, E. 1997. The expansion method, mathematical modelling and spatial econometrics, *International Regional Science Review*, vol. 20, pp. 9–32.

Chang, J & Hsieh, P. 2011. Particle swarm optimisation based on back propagation network forecasting exchange rates, I*nternational Journal of Innovative Computing, Information and Control*, vol. 7, no. 12 pp. 6837–6847.

Cliff, A.D. & Ord, J.K. 1981. Spatial processes: models and applications, Pion Limited, London.

Collobert, R & Bengio, S. 2004. Links between perceptrons, MLPs and SVMs, *Proceedings of the 21st International Conference on Machine Learning*, Page 23 ACM, New York, USA

Court, A.T. 1939. Hedonic price indexes with automobile examples, In General Motors Corporation, *The Dynamics of Automobile Demand*, New York, pp. 99–177.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*, John Wiley & Sons Inc.

Crooper, L.M, Deck, L.B. & McConnell, K.E. 1988. On the choice of functional form for hedonic functions, *The Review of Economics and Statistics*, vol. 70, no. 4, pp. 668–678.

Cui, D & Curry, D. 2005. Prediction in marketing using the support vector machine, *Marketing Science*, vol. 24, no. 4, pp. 595–615.

d' Amato, M & Kauko, T. 2008. Property market classification and mass appraisal methodology, In Kauko, T and d'Amato, M (Eds) *Mass Appraisal Methods: An International Perspective for Property Valuers*, pp. 280–303. Oxford: Wiley-Blackwell

de Smith M J (2011) *STATSREF: Statistical Analysis Handbook - A Web-based Statistics Resource.* The Winchelsea Press, Winchelsea, UK

Des Rosier, F & Thériault, M. 2008. Mass appraisal, hedonic price modelling and urban externalities: understanding property value shaping processes, in Kauko, T and d'Amato, M (Eds) *Mass Appraisal Methods: An International Perspective for Property Valuers*, pp. 1–24. Oxford: Wiley – Blackwell.

Diggle, P.J. & Ribeiro, P.J. 2007. *Model-based geographical*, New York, Springer.

Ding, Y.R., Cai, Y.J., Sun, P.D & Chen, B. 2014. The use of combined neural networks and genetic algorithms for prediction of river water quality, *Journal of Applied Research and Technology*, vol. 12, pp. 493–499.

Do, A.Q. & Grudnitski, G. 1992. A neural network approach to residential property appraisal, *The Real Estate Appraiser*, vol. 58, no. 3, pp. 38–45.

Dormann, C.F. *et al*. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review, *Ecography*, vol. 30, pp. 609–628.

Dubin, R., Pace, K.R. & Thibodeau, T.G. 1999. Spatial autoregression techniques for real estate data, *Journal of Real Estate Literature*, vol. 7, pp. 79–95.

Eckert, J.K. 1990. Property appraisal and assessment administration. International Association of Assessing Officers.

El Hamzaoui, Y & Perez, JAH. 2011. Application of artificial neural networks to predict the selling price in the real estate valuation process, *Proceedings of 10th Mexican International Conference on Artificial Intelligence*, Puebla, Mexico, Nov. 26-Dec. 4, pp. 75–181

Engelbrecht, A.P. 2007. *Computational Intelligence: An Introduction,* 2nd Edition, England, John Wiley & Sons, Limited.

Evans, A, James, H & Collins, A. 1992. Artificial neural networks: an application to residential valuation in the UK, *Journal of Property Valuation and Investment*, vol. 11, no. 2, pp. 195–204.

Farber, S & Yeates, M. 2006. A comparison of localised regression models in a hedonic house price context, *Canadian Journal of Regional Science*, vol. 29, no. 3, pp. 405–420.

Feng, Y & Jones, K. 2015. Comparing multilevel modelling and artificial neural networks in house price prediction. in *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM 2015): Proceedings of a meeting held 8-10 July 2015, Fuzhou, China.*. vol. 1, 15, Institute of Electrical and Electronics Engineers (IEEE), pp. 108–114.

Fotheringham, A.S., Crespo, R. & Yao, J. 2015. Exploring, modelling and predicting spatiotemporal variations in house prices, vol. 54, pp. 417–426.

Fraser, A.S. 1957a. Simulation of genetic systems by automatic digital computers I: Introduction, *Australian Journal of Biological Science*, vol. 10, pp. 484–491.

Fraser, A.S. 1957b. Simulation of genetic systems by automatic digital computers II: effect of linkage on rates of advance under selection, *Australian Journal of Biological Science*, vol. 10, pp. 492–499.

Friedman. J. & W. Stuetzel. 1981. Projection Pursuit Regression. *Journal of the American Statistical Association*, vol. 76, pp. 817–823.

Garro, BA & Vázquez, RA. 2015. Designing artificial neural networks using particle swarm optimization algorithms, *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 369298, 20 pages, doi:10.1155/2015/369298

Garson, G. D. 1991. Interpreting neural-network connection weights, Artificial Intelligence Expert, vol. 6, pp. 47–51.

Gencay, R. & Yang, X. 1996. A prediction comparison of residential housing prices by parametric versus semi-parametric conditional mean estimators, *Economics Letters*, vol. 52, pp. 129–135.

Gloudemans, R. J. 1990. Adjusting for time in computer-assisted mass appraisal, *Property Tax Journal*, Vol., pp. 83–99.

Gloudemans, R. J. 2002. Comparison of three residential regression models: additive, multiplicative, and nonlinear, *Assessment Journal*, vol. 9, no. 4, pp. 25–36.

Gonzalez, M.A.M., Soibelman, L., & Formoso, C. T. 2005. A new approach to spatial analysis in CAMA, *Property Management*, vol. 23, no. 5, pp. 312–327.

Gonzalez, S.A.M & Formoso, C.T. 2006. Mass appraisal with genetic fuzzy rule Based systems, *Property Management*, vol. 24, no. 1, pp. 20–30.

Goodman, A.C. 1978. Hedonic prices, price indices and housing markets, *Journal of Urban Economics*, vol. 5, pp. 471–484.

Goodman, A.C & Thibodeau, T.G. Housing market segmentation, *Journal of Housing Economics*, vol. 7, pp. 121–143.

Gill, P.E., Murray, W., & Wright, M.H. 1981. *Practical Optimisation*. Academic Press.

Greene, H.W. 2003. *Econometric Analysis*, Fifth Edition, Prentice Hall, Upper Saddle River, New Jersey.

Grover, R. 2016. Mass valuations, *Journal of Property Investment and Finance*, vol. 34, no. 4, pp. 191–204.

Guan, J, Zurada, J & Levitan, AS. 2008. An adaptive neuro-fuzzy inference system based approach to real estate property assessment. *Journal of Real Estate Research*, vol. 30, no. 4, pp. 395–420.

Gupta, J.N.D. & Sexton, R.S. 1999. Comparing backpropagation with a genetic algorithm for neural network training, *Omega*, vol. 27, pp. 679–684.

Haas, T.C. 1996. Multivariate spatial prediction in the presence of non-linear trend and covariance non-stationarity, *Environmetrics*, vol. 7, no. 2, pp. 145–165.

Hagan, M.T & Menhaj, M.B. 1994. Training feedforward networks with the Marquardt Algorithm, *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993.

Haining, R. 2003. *Spatial Data Analysis- Theory and Practice*, Cambridge, Cambridge University Press.

Han J., & Kamber M. 2006. *Data Mining: concepts and techniques*. Morgan Kaufmann Publishers.

Hastie, T. & Tibshirani, R. 1990. *Generalized Additive Models*. Chapman and Hall-New York

Hassan., R. Cohanim., B. de Weck., O & Venter, G. 2005. Comparison of particle swarm optimisation and the genetic algorithm, in *The Proceedings of 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, vol. 2, pp. 18–21.

Holland, J.H. 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press

Holmes, G, Hall, M & Frank, E. 1999. Generating rules from model trees, Working Paper Series 99, No.2, pp. 1–9. ISSN 1170 – 487X.

Hotta, K. 2008. Scene classification based on multi-resolution orientation histogram of Gabor features, in *Proceeding of the 6th International Conference on Computer Vision Systems*, vol. 5008 of Lecture Notes in Computer Science, pp. 291–301.

Hornik, K. 1991. Approximation capabilities of multilayer feed-forward networks, *Neural Networks*, vol. 4, pp. 251–257.

Huang, B., Wu, B., & Barry, M. 2010. Geographically and temporally weighted regression for modelling spatio-temporal variation in house prices, *International Journal of Geographical Information Science*, vol. 24, no. 3, pp. 383–401.

IAAO. 2013. *Standard on Mass Appraisal of Real Property*, access online on 4th February, 2015 via: http://katastar.rgz.gov.rs/masovna-procena/Files/2.Standard_of_Mass_Appraisal_of_Real_Property_2013.pdf

IAAO. 2014. *Guidance on International Mass Appraisal and Related Tax Policy*, Kansas City, IAAO.

Intrator, O & Intrator, N. 2001. Interpreting neural-network results: a simulation study, *Computational Statistics and Data Analysis*, vol. 37, no. 3, pp. 373–393.

Janssen, C. & Söderberg, B. 1999. Estimating market prices and assessed values for income properties, *Urban Studies*, vol. 36, no. 2, pp. 359–396.

Kang H. B. & Reichert, A.K. 1987. An Evaluation of Alternative Estimation Techniques and Functional Forms in Developing Statistical Appraisal Models‖, *Journal of Real Estate Research*, vol. 2, no.1, pp. 1–27.

Kauko, T & d'Amato, M. 2008. Introduction: Suitability issues in mass appraisal methodology, in Kauko, T and d'Amato, M (Eds) *Mass Appraisal Methods: An International Perspective for Property Valuers*, pp. 1–24. Oxford: Wiley – Blackwell.

Kauko, T. 2003. On current neural network application involving spatial modelling of property prices, *Journal of Housing and the Built Environment*, vol. 18, pp. 159–181.

Kathmann, RM. 1993. Neural networks for the mass appraisal of real estate, *Computer, Environment and Urban Systems*, vol. 17, issue 4, pp. 373–384.

Kaveh, A, Bakhshpoori, T & Ashoory, M. 2012. An efficient optimisation procedure based on cuckoo search algorithm for practical design of steel structures, *International Journal of Optimisation in Civil Engineering*, vol. 2, no. 1, pp. 1–14.

Keitt, T.H., Bjørnstad, O.N., Dixon, P.M. & Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions, *Ecography*, vol. 25, no 5, pp. 616–625.

Kennedy, J & Eberhart, R. 1995. Particle swarm optimisation, in *Proceeding of IEEE International Conference on Neural Networks*, Pert, Australia, Nov 27-Dec 1, pp. 1942–1948.

Kennedy, J. & Eberhart, R.C. 1997. A discrete binary version of the particle swarm algorithm. System, Man, and Cybernetics, in: *Proceedings of 1997 IEEE international conference on Computational Cybernetics and Simulation*, vol. 5, 12–15, October, pp. 4104–4108.

Kestens, Y., Thériault, M & Des Rosier, F. 2006. Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles, *Journal of Geographical Systems*, vol. 8, no. 1, pp. 61–96.

Kilpatrick, J. 2011. Expert systems and mass appraisal, *Journal of Property Investment and Finance,* vol. 29, nos. (4/5), pp. 529–550.

Kissling, W.D & Karl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecology and Biogeography,* vol. 17, pp. 59 – 71.

Kryvobokov, M. 2004. Urban land zoning for taxation purposes in Ukraine: possible methods under an immature land market, *Property Management*, vol. 22, nos. 3/4, pp. 214–229

Krause, A.L. & Bitter, C. 2012. Spatial econometrics, land values and sustainability: trends in real estate valuation research, *Cities*, vol. 29, pp. S19–S25.

Kwok, T.Y., & Yeung, D.Y. 1997. Constructive algorithms for structure learning in feedforward neural networks for regression problems, *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 630–645.

Lam, K.C., Yu, C.Y & Lam, K.Y. 2009. Support vector machine and entropy based decision support system for property valuation, *Journal of Property Research*, vol. 26, issue 3, pp. 213–233.

Lancaster, K. 1966. A new approach to consumer theory, *The Journal of Political Economy*, vol. 74, no. 2, pp. 23–29.

Lenk K, Worzala, E.M & Silva, A. 1997. High-tech valuation: should artificial neural networks bypass the human valuer? *Journal of Property Valuation and Investment*, vol. 15, pp. 8–26.

Levenberg, K. 1944. A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, vol. 5, pp. 164–168.

Liang, Y.H. 2008. Combining neural networks and genetic algorithms for predicting the reliability of repairable systems, *International Journal of Quality & Reliability Management*, vol. 25, no. 2, pp. 201–210.

Lichstein, J.W., Simons, T.R., Shriner, S.A., & Franzreb, K.E. 2002. Spatial autocorrelation and autoregressive models in ecology, *Ecological Monograph*, vol. 72, pp. 445–463.

Limsombunchai, V., Gan, C & Lee, M. 2004. House price prediction: hedonic price model vs artificial neural network. *American Journal of Applied Sciences*, vol. 1, no 3, pp. 193–201.

Lin, C. 2010. *Critical Analysis and Effectiveness of Key Parameters in Residential Property Valuations*, PhD Thesis University of Buffalo, State University of New York, U.S.A.

Lin, C.C. & Mohan, S.B. 2011. Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, vol. 4, no. 3, pp. 224–243.

Lv, J. Wang, X, Xie, D. & Wei, C. 2008. A hybrid approach of neural network with particle swarm optimization for tobacco pests prediction, In: Li, D. & Zhao, C. (Eds), *Computer and Computing Technologies in Agriculture* II, Springer, Beijing, China, vol. 2, Springer, pp. 1251–1260.

Malpezzi, S, Ozanne, L & Thibodeau, T. 1980. Characteristics prices of housing in fifty-nine metropolitan areas, Research Report, Washington, DC, The Urban Institute, December

Mark, J & Goldberg, M. 1988. Multiple regression analysis and mass assessment: a review of the issues, *Appraisal Journal*, vol. 56, no. 1, pp. 89–109.

Marquardt, D. 1963. An algorithm for least square estimation of nonlinear parameters, SIAM *Journal on Applied Mathematics*, vol. 11, no 2, pp. 431–441.

Masters, T. 1993. *Practical Neural Network Recipes in C++*, Academic Press, Boston, MA. McGraw Hill

Melanie, M. 1996. *An Introduction to Genetic Algorithms*. A Bradford Book, Massachusetts, the MIT Press.

McCluskey, W.M. 1996. Predictive accuracy of machine learning models for the mass appraisal of residential property, *New Zealand Valuers' Journal*, July, pp. 41–47.

McCluskey, W.J. 1997. A critical review of computer assisted mass appraisal techniques. In McCluskey, W.J. & Adair, A.S. (Eds.) *Computer Assisted Mass Appraisal: An International Review*, Ashgate, London, pp. 1–25.

McCluskey, W.J. 2016. Real property taxation in the republic of Kazakhstan, *Land Tenure Journal*, no. 2, pp. 119–138.

McCluskey, W., Deddis, W., Mannis, A., McBurney, D & Borst, R. 1997. Interactive application of computer assisted mass appraisal and geographic information systems, *Journal of Property Valuation and Investment*, vol. 15, no. 3, pp. 448–465.

McCluskey, W & Anand, S. 1999. The application of intelligent hybrid techniques for the mass appraisal of residential properties, *Journal of Property Investment and Finance*, vol. 17, no. 3, pp. 218–238.

McCluskey, W.J. & Borst, R.A. 2007. Specifying the effect of location in multivariate valuation models for residential properties, *Property management*, vol. 23, no. 4, pp. 313–343.

McCluskey, W.J. & Borst, R.A. 2011. Detecting and validating residential housing submarkets: a geostatistical approach for use in mass appraisal, *International Journal of Housing Market and Analysis*, vol. 4, no. 3, pp. 290–318.

McCluskey, W, Davis, P, Haran, M, McCord, M & McIlhatton, D. 2012. The potential of artificial neural networks in mass appraisal: the case revisited, *Journal of Financial Management of Property and Construction*, vol. 17, no. 3, pp. 274–292.

McCluskey, W.J., McCord, M, Davis, P.T, Haran, M & McIlhatton, D. 2013. Prediction accuracy in mass appraisal: a comparison of modern approaches, *Journal of Property Research*, vol. 30, no. 4, pp. 239–265.

McGreal, S, Adair, A, McBurney, D & Patterson, D. 1998. Neural networks: the prediction of residential values, *Journal of Property Valuation and Investment*, vol. 10, no. 1, pp. 57–70.

Michalewicz, Z. 1996. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin Heidelberg, New York, Third edition

Militino, A.F., Ugarte, M.D. & Garcia-Reinados, L. 2004. Alternative models for describing spatial dependence among dwelling selling prices, *Journal of Real estate Finance and Economics*, vol. 29, no. 2, pp. 193–209.

Mirjalili, S.A., Mohd, H.S.Z. & Sardroudi, H.M. 2012. Training feedforward neural networks using hybrid particle swarm optimisation and gravitational search algorithm, *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 11125–11137.

Møller, F.M. 1993. A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, vol. 6, pp. 525–533.

Mueller, J.M. & Loomis, J.B. 2008. Sptial dependence in hedonic property models: do different corrections for spatial dependence results in economically significant difference in estimated implicit prices? *Journal of Agricultural and Resource Economics*, vol. 33, no. 2, pp. 212–231.

Nawi, N.M, Khan, A & Rehman, M.Z. 2013a. A new back propagation neural network optimised with cuckoo search algorithm, *Computational Science and its Application-ICCSA*, vol. 7971, pp. 413–426.

Nawi, N.M., Khan, A & Rehman, M.Z. 2013b. CSLM: Levenberg-Marquardt based back propagation algorithm optimised with cuckoo search, *Journal of ICT Research and Applications*, vol. 7, no. 2, pp. 103–116.

Nguyen, N. & Cripps, A. 2001. Predicting housing values: a comparison of multiple regression analysis and artificial neural networks, *Journal of Real Estate Research*, vol. 22, no. 3, pp. 313–336.

Openshaw, S. 1998. Neural network, genetic, and fuzzy logic models of spatial interaction, *Environmental and Planning A*, vol. 30, pp. 1857–1872.

Orozco, J. & García, C. 2003. Detecting pathologies from infant cry applying scaled conjugate gradient neural networks. *Paper Presented at the European Symposium on Artificial Neural Networks*, Bruges (Belgium), 23–25 April.

Osland, L. 2010. An application of spatial econometrics in relation to hedonic house price modelling, *Journal of Real Estate Research*, vol. 32, no. 2, pp. 289–320.

Osman, H, Omar, S.S. & Mustapha, A.S. 2013. A machine learning for stock market prediction, *International Journal of Computer Science and Telecommunications*, vol. 4, issue 12, pp. 17–23.

Osuna, E., Freund, R., & Girosi, F. 1997. Training support vector machines: an application to face detection, *Proceedings of Computer Vision and Pattern Recognition*, June 17–19 Puerto Rico, pp. 130–136.

Pace, R.K & Gilley, O.W. 1997. Using the spatial configuration of the data to improve estimation, *Journal of Real Estate Finance and Economics*, vol. 14, no. 3, pp. 333–340.

Pace, R.K., Barry, R., Clapp, J.M & Rodriquez, M. 1998. Spatiotemporal autoregressive models of neighborhood effects, *Journal of Real Estate Finance and Economics*, Vol. 17:1, pp. 15–33.

Pace, R.K., Barry, R., Gilley, O.W. & Sirmans, C.F. 2000. A method for spatial-temporal forecasting with an application to real estate prices, International Journal of Forecasting, vol. 16, pp. 229–246.

Palmquist, R.B. 1980. Alternative techniques for developing real estate price indexes, The Review of Economics and Statistics, vol. 62, no. 3, pp. 442–448.

Papakostas, G., Boutalis, Y., Samartzidis, S., Karras, D. & Mertzios, B. 2005. Combining backpropagation and genetic algorithms to train neural networks. In the *12th Inernational Workshop on Systems, Signals & Image Processing, 22-24 September 2005, Chalkida, Greece*, pp. 171–177.

Peterson, S & Flanagan, A.B. 2009. Neural network hedonic pricing models in mass real estate appraisal, *Journal of Real Estate Research*, vol. 31, no 2, pp. 147–164.

Platt, J.C. 1998. Fast training of support vector machines using sequential minimal optimization, in *B Scolkopf, C Burges and A Simola (eds), Advances in Kernel Methods: Support Vector Machines*, Cambridge, MA: MIT Press.

Powell, M.J.D. 1977. Restart procedures for the conjugate gradient method, *Mathematical programming*, vol. 12, pp. 241–254.

Quilan, J.R. 1992. Learning with continuous classes, *Proceedings of 5th Australian Joint Conference on Artificial Intelligence*, Singapore, pp. 343–348.

Rangel, T.F.L.V.B, Field, R & Diniz-Filho, J.F.A. 2011. SAM Tutorial, https://www.ecoevol.ufg.br/sam/SAMTutorial.pdf

Reed, J., Toombs, R & Barricelli, N.A. 1967. Simulation of biological evolution and machine learning, *Journal of Theoretical Biology*, vol. 17, pp. 319–342.

Riedmiller, M & Braun, A. 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *International Conference on Neural Networks (ICNN-93*, San Francisco, CA), pp. 586–591. IEEE Press, Piscataway, New Jersey, USA

Rosen, S. 1974. Hedonic prices and implicit markets, product differentiation in pure competition, *Journal of Political Economy*, vol. 82, pp. 218–233.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. 1986. Learning representations by back-propagating errors, *Nature*, vol. 323, pp. 533–536.

Schulz, R, Wersing, M & Werwatz, A. 2014. Automated valuation modelling: a specification exercise, *Journal of Property Research*, vol. 31, no 2, pp. 131–153.

Sharma, A.K., Sharma, R.K. & Kasana, H.S. 2007. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modelling, *Applied Soft Computing*, vol. 7, issue 3, pp. 1112–1120.

Shin, K.Y., Lee, T.S. & Kim, H.J. 2005. An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, vol. 28, no. 1, pp. 127–135.

Sirmans, G.S, Macpherson, D.A. & Zietz, E.N. 2005. The composition of hedonic pricing models, *Journal of Real Estate Literature*, vol. 13, no. 1, pp. 3–43.

Stenehjem, E. 1974. Automated mass appraisal of real property – a scientific approach to the mass appraisal of residential property, *International Association of Assessing Officers*, pp. 22–43.

Sun, H., Tu, Y. & Yu, S. 2005. A spatial-temporal autoregressive model for multi-unit residential market analysis, *Journal of Real Estate Finance & Economics*, vol. 31, no. 2, pp. 155–187.

Suresh, A., Harish, K.V, & Radhika, N. 2015. Particle swarm optimisation over back propagation neural network for length of stay prediction, *Precedia Computer Science*, vol. 46, pp. 268–275.

Taherian, H. et al. 2013. Application of an improved neural network using cuckoo search algorithm in short-term electricity price forecasting under competitive power markets, *Journal of Operation and Automation in Power Engineering*, vol. 1, no. 2, Summer & Fall, pp. 136–146.

Tay, F.E.H., Cao, L. 2001. Application of support vector machines in financial time series forecasting, *Omega*, vol. 29, pp. 309–317.

Tay, D.P.H. & Ho, D.K.K. 1992. Artificial intelligence and the mass appraisal of residential apartment, *Journal of Property Valuation and Investment*, vol. 10, pp. 525–540.

Theriault, M. Des Rosier, F., Villeneuve, P. & Kestens, Y. 2003. Modelling interactions of location with specific value of housing attributes, *Property Management*, vol. 21, no. 1, pp. 25–48.

Thibodeau, T.G. 2003. Marking single-family property values to market, *Real Estate Economics*, vol. 31, no 1, pp. 1–22.

Thompson, J.F. 2008. Data issues involved with the application of automated valuation methods: a case study, In Kauko, T and d'Amato, M (Eds) *Mass Appraisal Methods: An International Perspective for Property Valuers*, pp. 27–48. Oxford: Wiley-Blackwell.

Tian, L. & Noore, A. 2005. Evolutionary neural network modelling for software cumulative failure time prediction, *Reliability Engineering and System Safety*, vol. 87, issue 1, pp. 45–51.

Tobler, W. 1979. Celluar Geography. In S. Gale and G. Olsson (eds.), *Philosophy in Geography*. Dordrecht.

Tu, Y, Sun, H & Yu, S. 2007. Spatial autocorrelations and urban housing market segmentation, *Journal of Real Estate finance & Economics*, Vol. 34, pp. 385–406.

Valente, J., Wu, S., Gelfand, A. & Sirmans, C.F. 2005. Apartment rent prediction using spatial modelling, *Journal of Real Estate Research*, vol. 27, no. 1, pp. 105–136.

Valian, E., Mohanna, S. & Tavakoli, S. 2011. Improved cuckoo search algorithm for feedforward neural network training, *International Journal of Artificial Intelligence & Applications*, vol.2, no.3, pp. 36–43.

Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*, Springer, New York.

Vapnik, V.N. 1999. An overview of statistical learning theory, *IEEE Transaction on Neural Networks*, vol. 10, no. 5, pp. 988–999.

Verkooijen, W.J.H. 1996. *Neural networks in economic modelling*, doctoral dissertation, Centre for Economic Research, Tilburg University, Tilburg, pp. 205.

Wall, M.M. 2004. A close look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning & Inference*, vol. 121, pp. 311–324.

Walton, S., Hassan, O, Morgan, K & Brown, MR. 2011. Modified cuckoo search: a new gradient free optimisation algorithm, *Journal of Chaos, Solitons and Fractals*, vol. 44, pp. 710–718.

Wang, Y & Witten, H.I. 1997. Inducing model trees for continuous classes, in *Proceedings of 9th European Conference on Machine Learning*, Prague, Czech Republic.

Werbos, P.J. 1974. Beyond regression: new tools for prediction and analysis in behavioral sciences. Doctoral Dissertation Applied Mathematics, Harvard University

Wilhelmsson, M. 2002. Spatial Models in Real Estate Economics, *Housing, Theory and Society*, vol. 19, pp. 92–101.

Wiering, M.A, van der Ree, M.H., Embrechts, M.J, Stollenga, M.F., Meijster, A, Nolte, A & Schomaker, LRB. 2013. The neural support vector machine, in *Proceedings of the 25th Benelux Artificial Intelligence Conference (BNAIC)*, November 7- 8, pp. 247–254.

Wiering, M.A. & Schomaker, L.R.B. 2014. Multilayer support vector machines, in *Johan A.K. Suykens, Marco Signoretto, Andreas Argyriou (Eds.) Regularization, Optimization,* *Kernels, and Support Vector Machines*, Edition: CRC Machine Learning and Pattern Recognition Series, Publisher: Chapman & Hall

Wilamowski, B.M., Kaynak, O, Iplikci, S & Efe, M.Ö. 2001. An algorithm for fast convergence in training neural networks, *IEEE*, pp. 1778–1782.

Wilamowski, B.M. & Yu, H. 2010. Improved computation for Levenberg-Marquardt Training, IEEE Transactions on Neural Networks, vol. 21, no. 6, pp. 930–937.

Worzala, E.M., Lenk, M.M. & Silva, A. 1995. An exploration of neural networks and its application to real estate valuation, *Journal of Real Estate Research*, vol. 10, no 2, pp. 185–202.

Yang, X. & Deb, S. 2009. Cuckoo search via levy flights, *Proceedings of World Congress on Nature and Biologically Inspired Computing*, India, pp. 210–214.

Yang, X-S. & Deb, S. 2010. Engineering optimisation by cuckoo search, *International Journal of Mathematical and Numerical Optimisation*, vol. 1, no. 4, pp. 330–343.

Yi, J.H., Xu, W.H. &Chen, Y.T. 2014. Novel back propagation optimization by cuckoo search algorithm, *The Scientific World Journal*, vol. 2014, Article ID 878262, 8 pages.

Yu, H & Wilamowski, B.M. 2011a. Levenberg-Marquardt training, in *the Industrial Electronics Handbook*, Vol. 5, Intelligent Systems, 2nd ed. (CRC Press, Boca Raton).

Yu, H. & Wilamowski, B.M. 2011b. Neural Network Training with Second Order Algorithms. In Hippe, Z.S. et al. (Eds), *Human – Computer Systems Interaction. Background and Applications*, 2nd Series: Advances in Intelligent and Soft Computing (99 Part II) Springer-Verlag, pp. 463 – 476.

Zhang, J, Zhang, J, Lok, T, & Lyu, M.R. 2007. A hybrid particle swarm optimisation-back propagation algorithm for feedforward neural network training, *Applied Mathematics and Computation*, vol. 185, pp. 1026–1037.

Zurada, J, Levitan, A.S. & Guan, J. 2011. A comparison of regression and artificial intelligence methods in a mass appraisal context, *Journal of Real Estate Research*, vol. 33, no. 3, pp. 349–387.