USING GEOGRAPHICAL INFORMATION SYSTEM - MULTIPLE
REGRESSION ANALYSIS - GENERATED LOCATION VALUE RESPONSE
SURFACE APPROACH TO MODEL LOCATIONAL FACTOR IN THE
PREDICTION OF RESIDENTIAL PROPERTY VALUES

CHIN CHUI VUI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Science (Real Estate)

Faculty of Geoinformation Science and Engineering
Universiti Teknologi Malaysia

JANUARY 2006
To my believable god
ACKNOWLEDGEMENTS

First of all, my deep appreciation and gratitude dedicate to my supervisors, Prof. Madya Dr. Abdul Hamid bin Hj. Mar Iman and Prof. Madya Dr. Buang bin Alias for their patient and excellent guidance with constructive comments on the direction during the preparation of this thesis.

I am grateful and feel indebted to Prof. Madya Dr. Taher bin Buyong at the department of Geoinformatic for his generosity and willing to share attitude to share with me their constructive and valuable comments and experiences throughout the research period.

Not forgetting research assistants in the department of Geoinformatic, Eu Yen Zhing, Karen Lam, and Chai Beng Chung and research assistants in the department of property management, Nik Adib Bin Nik Din, Bismawaty Binti Hassan, and Zariman Bin Ibrahim, for their helpfulness in always share their Geographical Information System (GIS) skills during the research period. Staffs of Johor Bahru Valuation and Property Services Office (JPPh) are not being forgotten also for providing me the necessary information for this study.

Last but not least to my family, especially to my dad Chin Vui Ming and mom Shoon Kui Ying, and Chung Hwee for their never ending support and encouragement in every aspect until this thesis is completed.
ABSTRACT

The limitation and highly complex process of discrete measurement of location have encouraged searching for alternative approach to derive locational compensation factor in the prediction of property values. A new approach by use of Geographical Information System-Multiple Regression Analysis-generated location value response surface (GIS-MRA-generated LVRS) approach was proposed in this study. The method used LVRS in a GIS to model locational influence from MRA-generated residuals, with no locational variables on residential property values in the local context by integrating spatial & aspatial data in term of developing a hybrid predictive model. This study has three main objectives. First, to discuss the pertinent factors influencing residential property values, including the location factors. Second, to develop a predictive model of residential property values, whereby the locational factor is modelled by GIS-MRA-generated LVRS. Third, to examine the usefulness of GIS-MRA-generated LVRS to improve the quality of the regression model. To achieve these objectives, this study was divided into two main parts. The first part comprised the theories of value, residential property value factors, location modelling, MRA modelling, and the spatial interpolation techniques. The second part, comprised the development of hybrid models, whereby the locational factors was modelled by GIS-MRA-generated LVRS approach and examining its ability to improve regression modelling. As many as 125 individual terraced units in three adjoining housing schemes (Taman Pelangi, Taman Sentosa and Taman Sri Tebrau) in Johor Bahru were used for model estimation, while 14 transacted units were set aside for predictive purposes. Results have shown models applying LVRS have managed to improve overall model’s statistical quality and predictive performance by achieving higher proportion of “reasonably accurate” prediction as compared to the traditional MRA models. The LVRS has allowed a clearer spatial visual picture of the location influence to be captured at all level in the study area. The location factor influence to property value can be modelled in a more effective way.
ABSTRAK

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
<td>i</td>
<td></td>
</tr>
<tr>
<td>DECLARATION</td>
<td>ii</td>
<td></td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v – vi</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii - xiii</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiv - xv</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvii - xviii</td>
<td></td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xix</td>
<td></td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xx</td>
<td></td>
</tr>
</tbody>
</table>

1 INTRODUCTION

1.1 Introduction 1
1.2 Problem Statement 5
1.3 Objectives of Study 6
1.4 Scope of Study 6
1.5 Importance of Study 7
1.6 Methodology 7
1.6.1 Literature Review 8
1.6.2 Data Collection 8
1.6.3 Analysis 8
2 THEORY OF PROPERTY VALUES AND VALUE FACTORS

2.1 The Concept of Value 14
2.2 Market Value 16
2.3 Factors Influencing on Residential Property Values 17
  2.3.1 Market Factors 18
    2.3.1.1 Population 19
    2.3.1.2 Purchasing Power 19
    2.3.1.3 Inflation 21
    2.3.1.4 Government Policy 22
    2.3.1.5 Cost of Production 22
  2.3.2 Psychical Factors 22
    2.3.2.1 Structural Improvement and Materials Used 23
    2.3.2.2 Accommodation and Size 23
    2.3.2.3 Age and Condition of Repair 23
  2.3.3 Locational Factors 24
    2.3.3.1 Neighbourhood 25
    2.3.3.2 Site 29
  2.3.4 Other Factors 30
    2.3.4.1 Tenure Status 30
    2.3.4.2 Type of Ownership 30
2.4 Appropriate Variables Selection
   – Summary of the Previous Modelling 31
2.5 Conclusion 34

3 MODELLING OF LOCATION OF RESIDENTIAL PROPERTY VALUES AND MULTIPLE REGRESSION ANALYSIS (MRA)

3.1 Previous Studies on Modelling of Location - Traditional Approach 36
   3.1.1 Accessibility Factors 36
   3.1.2 Environmental Factors 37
   3.1.3 Neighbourhood or Sub-Market 37
3.2 Geo-Statistical Spatial Approaches – GIS-MRA-Generated LVRS 42
3.3 The Theory of MRA 44
   3.3.1 Error Term in Regression 46
   3.3.2 Basic Assumptions 48
   3.3.3 Appropriate Variable Selection 49
   3.3.4 Choosing the Proper Functional Form 49
   3.3.5 Diagnostics Testing 51
      3.3.5.1 Multicollinearity 52
      3.3.5.2 Misspecification 53
      3.3.5.3 Autocorrelation 53
      3.3.5.4 Heteroscedasticity 54
3.4 Conclusion 55
4 SPATIAL INTERPOLATION TECHNIQUES FOR CREATING LOCATION VALUE RESPONSE SURFACE (LVRS)

4.1 Introduction 57

4.2 Spatial Interpolation in GIS: Rationale for Locational Factor in Property Valuation 57

4.3 Interpolation Techniques 58

4.3.1 Global Surface Interpolation 59

4.3.1.1 Trend Surface 59

4.3.1.2 Fourier Series 60

4.3.2 Local Interpolators 60

4.3.2.1 Nearest neighbours 60

4.3.2.2 Spline 61

4.3.2.3 Inverse Distance Weighted (IDW) 62

4.3.3 Geostatistical Methods 64

4.3.3.1 Kriging 65

4.3.3.2 Type of Kriging 66

4.4 The Methodology of Ordinary Kriging 68

4.4.1 Modelling the Experimental Semivariogram 70

4.4.2 Model Fitting 73

4.4.2.1 Spherical model 74

4.4.2.2 Exponential model 75

4.4.2.3 Gaussian model 76

4.4.2.4 Linear model 77

4.5 Conclusion 79
## DATA AND ANALYSIS PROCEDURE

5.1 The Study Area .......................... 81
5.2 Data .............................. 82
   5.2.1 Property Sales Data .......................... 82
   5.2.2 GIS Data and Software ...................... 85
      5.2.2.1 Construction of the house lot digital map 85
      5.2.2.2 Spatial Interpolation Analysis 87
5.3 GIS-MRA-Generated LVRS Approach in Modelling the Value of Location 87
   5.3.1 Preliminary Model Specification ............. 87
      5.3.1.1 Test for Functional Forms 90
      5.3.1.2 Diagnostic Tests 90
   5.3.2 Location Adjustment ....................... 91
      5.3.2.1 GIS-MRA-Generated LVRS 91
      5.3.2.2 Construction of LVRS 91
5.4 Hybrid Model Quality Evaluation and Traditional Approach Used for Comparison 92
   5.4.1 Model Statistical Evaluation .................. 93
   5.4.2 Model Predictive Performance Evaluation 93
5.5 Conclusion ................................ 94

## RESULTS AND ANALYSIS

6.1 Basic Results .............................. 95
   6.1.1 Tests for Functional Form ................. 95
6.1.2 Regression Results for the Preliminary Model 96
6.1.3 Diagnostic Tests 97

6.2 LVRS 99

6.2.1 Application of Kriging 99
  6.2.1.1 Distribution of the Data Set 99
  6.2.1.2 Spatial Structure of Residual Values in the Study Area 101
  6.2.1.3 Searching Neighbourhood 104
  6.2.1.4 Cross-validation 105
  6.2.1.5 Prediction Standard Errors 106

6.2.2 LVRS Generated Using IDW and Kriging 107

6.3 Results of Hybrid Model 110
  6.3.1 Diagnostic Testing for the Hybrid Model 113

6.4 Hybrid Model Quality Evaluation 114
  6.4.1 Statistical Performance 115
  6.4.2 Predictive Performance 116

6.5 Conclusion 117

7 CONCLUSION AND RECOMMENDATION

7.1 Summary of the Study 118

7.2 Main Findings 120
  7.2.1 Factors Influencing Residential Properties Value in the Study Area – First Objective 120

7.2.2 Hybrid Model Developed Using GIS-MRA-Generated LVRS Approach – Second Objective 120
7.2.3 Improvement of Hybrid Model Using GIS-MRA-Generated LVRS Approach – Third Objective 121

7.3 Limitations of the Study 122

7.4 Recommendation for Future Research 123

REFERENCES 124

Appendices A – C 135 - 139
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Results of different models on terraced house in different location level</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>Independent variables included in the previous regression models</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Summary of variables used in previous modelling in residential properties</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>Traditional locational proxy variables used in the previous regression models</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Special cases of Box-Cox functional forms</td>
<td>51</td>
</tr>
<tr>
<td>5.1</td>
<td>Descriptive statistics of the property sample</td>
<td>84</td>
</tr>
<tr>
<td>5.2</td>
<td>List of variables used in the preliminary model and the unit of measurement</td>
<td>88</td>
</tr>
<tr>
<td>6.1</td>
<td>Adjusted $R^2$ from Box-Cox transformation on the basic model</td>
<td>95</td>
</tr>
<tr>
<td>6.2</td>
<td>Results for preliminary residential property value model</td>
<td>96</td>
</tr>
<tr>
<td>6.3</td>
<td>Partial correlation coefficient</td>
<td>98</td>
</tr>
<tr>
<td>6.4</td>
<td>Breusch-Pagan test, Ramsey’s RESET test and Durbin-Watson test for the base model</td>
<td>98</td>
</tr>
<tr>
<td>6.5</td>
<td>Calculated statistic from cross-validation for spherical and exponential model</td>
<td>106</td>
</tr>
<tr>
<td>6.6</td>
<td>Regression results which include location component</td>
<td>111</td>
</tr>
</tbody>
</table>
6.7 Breusch-Pagan test, Ramsey’s RESET test and Durbin-Watson test for the hybrid models 113

6.8 Regression results of the GIS-MRA-generated LVRS models and the traditional models 114

6.9 Predictive capability of the models 116

7.1 Regression results which include location component 120
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Flow chart of the research process</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>Various value factors of residential properties</td>
<td>18</td>
</tr>
<tr>
<td>3.1</td>
<td>Procedural steps for locational factor modelling used GIS-MRA-generated LVRS</td>
<td>44</td>
</tr>
<tr>
<td>3.2</td>
<td>Heteroscedasticity (Kelejian and Oates, 1974)</td>
<td>55</td>
</tr>
<tr>
<td>4.1</td>
<td>Inverse distance method for site sample analysis (Eric. and Pai, 2003)</td>
<td>62</td>
</tr>
<tr>
<td>4.2</td>
<td>Experimental semivariogram (Burrough and McDonnell, 1998)</td>
<td>72</td>
</tr>
<tr>
<td>4.3</td>
<td>Spherical model (Burrough and McDonnell, 1998)</td>
<td>75</td>
</tr>
<tr>
<td>4.4</td>
<td>Exponential model (Burrough and McDonnell, 1998)</td>
<td>76</td>
</tr>
<tr>
<td>4.5</td>
<td>Gaussian model (Burrough and McDonnell, 1998)</td>
<td>77</td>
</tr>
<tr>
<td>4.6</td>
<td>Linear model (Burrough and McDonnell, 1998)</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Location map of study area</td>
<td>82</td>
</tr>
<tr>
<td>5.2</td>
<td>House lot digital map and samples of house lot (house price 2001/03)</td>
<td>86</td>
</tr>
<tr>
<td>5.3</td>
<td>Process of digital map creation</td>
<td>86</td>
</tr>
<tr>
<td>6.1</td>
<td>The histogram of the residual values distribution</td>
<td>100</td>
</tr>
<tr>
<td>6.2</td>
<td>Cumulative frequency distribution of residual values</td>
<td>100</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.3</td>
<td>Semivariogram surface of residual values</td>
<td>101</td>
</tr>
<tr>
<td>6.4</td>
<td>Isotropic semivariogram model of residual values</td>
<td>102</td>
</tr>
<tr>
<td>6.5</td>
<td>Anisotropic semivariogram for the residual values points</td>
<td>104</td>
</tr>
<tr>
<td>6.6</td>
<td>Searching neighbourhood</td>
<td>105</td>
</tr>
<tr>
<td>6.7</td>
<td>Prediction standard errors map</td>
<td>107</td>
</tr>
<tr>
<td>6.8</td>
<td>IDW-based 2-D view of LVRS</td>
<td>108</td>
</tr>
<tr>
<td>6.9</td>
<td>Kriging-based 2-D view of LVRS</td>
<td>108</td>
</tr>
<tr>
<td>6.10</td>
<td>IDW-based 2 1/2-D view of LVRS</td>
<td>109</td>
</tr>
<tr>
<td>6.11</td>
<td>Kriging-based 2 1/2-D view of LVRS</td>
<td>109</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3BEDR</td>
<td>3 bedrooms</td>
</tr>
<tr>
<td>AA</td>
<td>Ancillary area</td>
</tr>
<tr>
<td>AGE</td>
<td>Age of building</td>
</tr>
<tr>
<td>CBD</td>
<td>Center Business District</td>
</tr>
<tr>
<td>FFINISH2</td>
<td>Class 2 floor finishes</td>
</tr>
<tr>
<td>GCOND</td>
<td>Good in condition</td>
</tr>
<tr>
<td>GFA</td>
<td>Gross floor area</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographical Information System</td>
</tr>
<tr>
<td>IDW</td>
<td>Inverse distance weighted</td>
</tr>
<tr>
<td>JPPH</td>
<td>Valuation and Property Service Office</td>
</tr>
<tr>
<td>JUPEM</td>
<td>Jabatan Ukur dan Pemetaan Malaysia</td>
</tr>
<tr>
<td>LA</td>
<td>Land area</td>
</tr>
<tr>
<td>LG</td>
<td>Log</td>
</tr>
<tr>
<td>LVRS</td>
<td>Location value response surface</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean absolute percentage error</td>
</tr>
<tr>
<td>MRA</td>
<td>Multiple Regression Analysis</td>
</tr>
<tr>
<td>PRIC</td>
<td>House price</td>
</tr>
<tr>
<td>SEE</td>
<td>Standard error of estimate</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Science</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum squared errors</td>
</tr>
<tr>
<td>T1</td>
<td>Single storey terrace</td>
</tr>
</tbody>
</table>
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ramsey’s Test</td>
<td>135</td>
</tr>
<tr>
<td>B</td>
<td>Durbin-Watson Test</td>
<td>136</td>
</tr>
<tr>
<td>C</td>
<td>Breusch-Pagan Test</td>
<td>138</td>
</tr>
</tbody>
</table>
1.1 Introduction

Residential property is a multi-dimensional heterogeneous commodity, characterized by its durability and structural inflexibility as well as spatial immobility. Past studies have advocated that each residential unit has a unique bundle of attributes such as accessibility to work, transport, amenities, the structural characteristics, neighbourhood, and environmental quality (Muth, 1960; Ridker and Henning, 1967; Stegman, 1969; Kain and Quigley, 1970; Evans, 1973; Lerman, 1979; So et al., 1997). In particular, these studies have indicated that house price function relates to location. It is commonly accepted that properties are spatially unique and this means that the location is an intrinsic attribute of a dwelling that will directly determine one housing quality and its market value. The importance of location in real estate is widely acknowledged and, as such, it is arguably the most important factor affecting property values (Ali, 2004).

However, modelling the location factor in property valuation has proved difficult because of the wide range of spatially defined attributes, which may or may not affect the value at a particular time and location. Furthermore, there is little consensus in the literature as to the best proxy for location factors, their measurements and how they influence property values. Consequently, modelling location for valuation purposes can be very difficult and subjective.
MRA is considered as a primary technique of the mass valuation models to explain and predict property values. In this context, MRA has been used to estimate residential property values in the U.S, since the 1950s and in U.K since the 1980s (Pendleton, 1965; Greaves, 1984; Adair and McGreal, 1988). This is also applied in other countries such as Australia, New Zealand, and Singapore, but has yet to be adopted in Malaysia. In applying MRA in property valuation, researcher must attempt to identify the data to be collected and variables to be measured in quantitative form. This task becomes more complicated when the location influence on residential property values needs to be identified.

Research that has sought to assess the determinants of property value has either ignored detailed location analysis or just deal with it only in a very general sense (Wyatt, 1997; So et al., 1997). To take an easy way out, some researchers simply omit the locational variables in their valuation models (Ferri, 1977). An interview has been conducted with local valuation-based firms and government offices such as Rahim & Co, Jurunilai Bersekutu. C.H.William & Talhar, Raja Hamzah, Ismail & Co, Zaki & Partners, T.D.Aziz, and Valuation and Property Services Offices (JPPH). Results found that the valuers infer a substantial amount of information about a property from its location, based on their local knowledge and experience. These have formed a variability of opinion among the valuers regarding the specific influence of location to property value. It is inevitable that no two valuers will apply same steps of action towards an opinion on location influence to a property value.

Location is an amalgamation of several factors that include a number of sources such as accessibility to shopping, employment, educational and leisure facilities; exposure to adverse environmental factors, such as traffic noise and hazard; neighbourhood amenity; perceived levels of neighbourhood security; and so on (Gallimore et al., 1996; McCluskey et al., 2000). From these, two key components of location can be isolated, i.e. neighbourhood quality and accessibility (McCluskey et al., 2000). Few, however, are capable of numeric measurement, but the measures may not always be valid representation of the influence, especially because of the complex interaction of value factors. For example, a property close to
excellent communication links may have a great influence on price of the property. The property value may also be reduced due to the impact of adverse environmental effects including busy traffic noise etc. It is apparent that all these factors will determine the particular property and housing quality. Therefore, by allowing one of these correlations have a substantial different impact in the estimated model.

Take for instance accessibility, the common approach applied in examining the locational factor is to include a distance variable from the Central Business District (CBD), which simply assumes that the location is homocentric. This is based on the traditional location theory that examines the role of accessibility to central locations on house price. However, it is argued that house prices are determined not only by accessibility but also by the environmental attributes of the location (Stegman, 1969; Richardson, 1971; Henderson, 1977; Pollakowski, 1982). Besides, there are also theories of multiple nuclei model incorporating the concentric pattern that are more appropriate for analyzing locational influence on property values. However, there are also some researchers who employ more sophisticated measurements of location in this aspect such as using the type of transport, time taken per trip, and transportation cost.

Apart from that, there is an approach that adjusts for location by partitioning the study area into neighbourhoods and each neighbourhood will be analyzed separately or categorizes each as a dummy variable (Azhari, 2001; Hamid, 2003). From the mass appraisal modelling perspective, it is essential to subdivide the study area into “realistic” sub-market or neighbourhoods to enable the model to reflect the influence of location more accurately. However, this could pose a modelling constraint in terms of data representativeness such as neighbourhoods with too few transactions this may give rise to small sample size in some neighbourhoods, which may not work well with the MRA technique in the statistical estimation. Too small sample size, therefore, is less concerned with explicit consideration of the values ascribable to the characteristics.
In addition, the problem commonly faced in the use of neighbourhood is the requirement for subjective judgments about the boundaries of each neighbourhood and the numeric indicator for neighbourhood quality. To solve this problem, some researchers have simply asked local valuers or local experts to rank the neighbourhood quality (Hickman et al., 1984). There is little consensus, however, on which variables are the best proxy for neighbourhood quality measurement. There is no exact answer whether it should be based on actual house price or property physical characteristic or housing quality or ward boundary or defined in spatial terms (Can, 1990; Adair et al., 1996). Therefore, neighbourhood quality is arguably an unobservable variable (Dubin and Sung, 1987). Either one of the proxies for neighbourhood quality measurement is adopted, it may lead to disparities or inconsistencies for properties adjoin or close to neighbourhood boundaries. A hard edge may be implied at such boundaries, whereas in reality the varying influence of location may operate far more smoothly and the spatial trends occur as opposed to distinct areas of homogeneous property subsets (Mackmin, 1989; Gallimore et al., 1996).

The complexity of locational factors and the problems to identify them, which confront their assessment, can seriously threaten the validity of a MRA model. This is due to the selection of location characteristics has an impact for the model estimation (Can and Megbolugbe, 1997; Raymond Tse, 2002). If the method of computing location characteristics may be inadequate, this will increase the error of estimation and reduce the predictive capability of the model. In fact, location is only one of many variables in the equation, and it is commonly agreed as an important variable influencing residential property values. For this reason, it appears in almost every house price regression. Hence, an approach, which accurately accommodates the transitions of locational influence (capturing the value of neighbourhood and accessibility) across a particular area from which valuation data are derived, is required.

This study uses a LVRS generated by using GIS in the modelling of locational influence on residential property values in the local context by integrating spatial and aspatial data taking Johor Bahru as a study area. In particular, this study
develops locational adjustment factor based on the residuals generated from location-blind model. The residuals or the discrepancies between actual and estimated prices can be regarded as the value of location and used to construct a LVRS. This surface could then be used to adjust for under-or over-valuation of any property within the area to estimate location influences. In this study, GIS will be used to generate the LVRS. In constructing LVRS, appropriate spatial interpolation analysis technique(s) within a GIS will be chosen for this purpose. In order to consider GIS-MRA-generated LVRS, this study will concentrate on the model quality improvement and predictive performance of the hybrid model (marriage between MRA model and location adjustment factor generated from the GIS-generated LVRS) in relation to predicting unsold properties whose sales prices are unknown. Also, the results of the traditional approaches in modelling the location effect will be used to compare model’s predictive capability.

1.2 Problem Statement

Having discussed the background scenarios, the issues of this study can be stated as follow:

1. What are the pertinent factors influencing residential property values?

2. How can the locational influence on the residential property values be quantified in the MRA model?

3. Will the locational effect modelled by the GIS-MRA-generated LVRS improve the regression results?
1.3 Objectives of Study

In line with the issues mentioned above, this study has the following objectives:

1. To discuss the pertinent factors influencing residential property values, including the location factors.

2. To develop a predictive model of residential property values using MRA, whereby the locational factors is modelled by GIS-MRA-generated LVRS.

3. To examine whether the use of GIS-MRA-generated LVRS can improve the regression results.

1.4 Scope of Study

There is a wide range of location attributes. Real estate is spatially unique in which location is an intrinsic attribute that directly determines the quality and market value of the property. Literature review reported that the modelling of the locational factors in property valuation is very complex because of the wide range of spatially attributes, which may or may not affect value at a particular time and location. There have also a little consensus as the best proxy for locational factor measurement. Besides, the combination of individual property-specific location variables is necessary because normally there are very few cases in each individual variable to be significant in the model (O’Conner, 2002). Therefore, this study is focusing on quantifying locational value factors in the model and considers location in a very general manner. In other words, the locational detail such as distance from school, leisure facilities, market and so on are not going to examine in this study.
1.5 Importance of Study

To adjust the property prices for locational influence, adoption of GIS-assisted approach can be used to enhance the traditional comparison method in property valuation, especially if mass appraisal approach has been adopted. The LVRS will tell something about locational differences in residential property values across a particular geographic area.

This can be used as a starting point to examine the cause-and-effect of locational differences in property values in the actual real estate market. This may help further understanding about the actual market forces that occur spatially. For example, in the low-residual areas, further investigation can be carried out to ascertain the actual factors giving rise to under-valuation. In the same way, in the high-residential areas, further investigation can be carried out to ascertain the factors that cause over-valuation.

It may be discovered that some in-situ factors have not been taken into account in the MRA but have actually determined the prices paid for the properties involved. For examples, unpleasant odours, streets pattern, topography, attitudes of residence to maintain the neighbourhood as a good place to live and so on. These are not capable using numeric measurement and even this is so, the measurements may not always be valid representing the influence.

1.6 Methodology

Briefly, the methodology of this study consists of three main components as follows. Further refinement to the procedure of this study is illustrated in Figure 1.1.
1.6.1 Literature Review

The literature review is spread over the next three chapters. Chapter 2 reviews the issues related to the theories of value and factors influencing residential property values, whereby the locational factors will be specifically emphasized. These involve wide spectrums of locational issues such as neighbourhood, accessibility, environmental quality, etc. Modelling the value of location, particular using the GIS-MRA-generated LVRS is presented in the following this entails, among other things, the theoretical foundation of the multiple regression models and its application. The literature on spatial interpolation analysis techniques is presented in Chapter 4.

1.6.2 Data Collection

There are two major categories of data needed in this study. They are valuation data and GIS data. The main source of these valuation data is obtained from the Johor Bahru Valuation and Property Services Office (JPPH). The value factors selected to build the statistical models in this study are primarily based on the availability of data from this department. The second source is GIS data. The analogue maps (Standard Sheet plans) are obtained from Jabatan Ukur dan Pemetaan Malaysia (JUPEM). The maps are constructed into digital form to provide the spatial information that is used in the GIS-MRA generation of value residuals at each geo-referenced location of the property in the study area.

1.6.3 Analysis

There are three major interrelated components of analysis as follows:
1.6.3.1 MRA Modelling

The first step in building the regression models is selecting the variables based on the literature review. The second step is specification of the appropriate functional forms of the regression models. Then, the regression results are tested to analyze their statistical qualities. The basic theories of property values and the property value factors are discussed in Chapter 2. The modelling aspects, and in particular the modelling of locational variables are discussed in Chapter 3. Data procedure is discussed in Chapter 5, and finally, the regression results are discussed in Chapter 6.

1.6.3.2 Spatial Interpolation Analysis

This analysis is carried to construct the LVRS. As a note, \( R = v - \hat{v} \), where \( \hat{v} \) is the estimated price from the regression without locational components; \( v \) is the actual price of the residential property; and \( R \) is the discrepancies between the actual and predicted price (residual or error of prediction), which capture locational influence. This discrepancies value will be fitted into each locational point to build a LVRS. Therefore, this response surface could then be used to adjust for under- or over-valuation (discrepancies) of properties in the study area. This enables locational influence to be measured and accurately accommodated into the transitions of locational factors effect across the study area. The locational influence can be measured for any point of property site within the area in relation to predict property price. The LVRS is generated by using spatial interpolation analysis techniques available within GIS. The basic theories and the rationalism of interpolation techniques used in this study are discussed in Chapter 4.

1.6.3.3 Regression Models Evaluation

The ability of the GIS-MRA-generated LVRS to improve the predictive model is evaluated on the basis of quality and predictive performance of the other models. Different spatial interpolation analysis techniques (IDW and kriging) for
generating LVRS are tested. Traditional locational factors calibration techniques are also compared. Model quality is evaluated based on their basic statistical significance while the predictive performance of the models is evaluated based on the size of proportion of mean absolute percentage error (MAPE) of each model and the proportion of accurate predictions from the holdout sample. These are discussed in Chapter 5 and Chapter 6.
**Issue:** The limitation and highly complex process for discrete measurement of location have encouraged searching for alternative approach to derive locational compensation factor. New approach should be based on GIS-MRA-generated LVRS

**Literatures on value factors include locational factors**

**Data collection**

Property sales data:
- Price
- Lot number
- Land Area
- Gross Floor Area
- Ancillary Area
- Type of lot
- Age
- Date of transaction
- Property types
- Interest in property
- Nature of transaction
- Type of fencing
- Type of finishes
- No. of bathroom/toilet
- Physical improvement
- Kitchen cabinet
- Kitchen extension
- Building extension
- No. of bedrooms
- State of repair

GIS data:
- Digitize Standard Sheet plans for study area

**OBJECTIVE 1**

**OBJECTIVE 2**

**GIS-MRA-generated LVRS**

\[ \hat{v} = f(X_1, X_2, X_3, \ldots, X_n) \]

\[ v - \hat{v} = R \]

\[ R = L \]

**OBJECTIVE 3**

**Model evaluation**

*Improvement of model quality:*
- Statistical result

*Predictive Performance:*
- Holdout samples of property sales data
  - MAPE
  - Proportion of accurate prediction

**Figure 1.1:** Flow chart of the research process

Where,

\[ V \] - property values; \( \hat{V} \) - predicted property values; \( X_{1,n} \) - factors beside locational factors; \( R \) - residual; \( L \) - locational factors
1.7 Chapters Layout

The contents of this study are divided into five chapters:

Chapter 1 – This chapter introduces the background of the study; problem statement; objectives; importance of the study; scope of study; methodology and chapter layout.

Chapter 2 – This chapter discusses the theories of property values and factors influencing residential property values.

Chapter 3 – This chapter contains an overview of previous studies in the modelling of locational factors for property valuation. These can be divided into two sub-categories of models. They are traditional statistical approach and hybrid approach (modelling the value of location using the GIS-MRA-generated LVRS). This chapter includes the discussion on the theoretical foundation of the multiple regression models and its application.

Chapter 4 – This chapter presents the spatial interpolation analysis techniques used to estimate the configuration of LVRS. The first part presents an overview of the interpolation techniques in GIS. The second part discusses the two interpolation techniques selected for this study.

Chapter 5 – This chapter explains the data and research procedure. The first part discusses the data used in this study. The second part explains how to model locational effect by using GIS-MRA-generated LVRS. The third part demonstrates how the spatial interpolation analysis techniques (IDW and kriging) within a GIS are applied to generate a response surface. Then, the analysis of model improvement follows.

Chapter 6 – This chapter presents the results of the study. The first part includes the discussion on the MRA and the LVRS of a hybrid model. In the second part, the
statistical significance and predictive performance of the hybrid models are compared with the traditional MRA models.

Chapter 7 – This chapter concludes the study as a whole and gives some recommendations for further study.


