GIS-MRA Techniques in Property Valuation: A Framework for Implementation

Abdul Hamid Mar Iman
Department of Property Management
Universiti Teknologi Malaysia, Skudai.
e-mail: hamid@fksg.utm.my

Abstract

The awareness among valuers to improve valuation analyses has induced the use of Geographic Information System (GIS) as a spatial tool. This tool has, at least, two main attractions to property valuation. First, the capability of providing an integrated database handling system, simultaneously relating spatial and attribute data of properties. Second, the capability of spatial operation and representation of these data. These functions can be performed in conjunction with multiple regression analysis (MRA) for a number of spatial analyses to improve modelling analyses. The MRA has been used as an aspatial analysis tool for real estate problems since 1920’s. Characterized by its “correlation” and “prediction” features, this technique provides statistical reasoning capabilities. However, although the statistical milieu provides models that are capable of producing relatively accurate and consistent regression estimates, the technique does not possess spatial analysis capability. This capability is felt necessary because the incidence of property values are locationally distributed. Furthermore, many complex phenomena that change spatially and temporally can only be better understood by analyzing them in a visualized manner. GIS fills in this gap. With a proper framework, both GIS and MRA can be combined as a new tool of spatial reasoning in real estate analyses, particularly valuation. The core of GIS-MRA techniques is spatial-statistical analyses of property determinants and how they affect property values. This paper presents an outline of the major aspects of GIS-MRA applications in property valuation with the objective of providing a basic theoretical framework of GIS-MRA techniques, focusing on the major steps of applications in property valuation. From the discussion, it is concluded that, if the GIS-MRA techniques are carefully employed, they will add a scientific dimension to real estate analyses.

Keywords: Geographic information system, multiple regression analysis, property valuation.

1.0 INTRODUCTION

The use of multiple regression analysis (MRA) in conjunction with Geographic Information System (GIS) is only a recent development in property valuation. Simply, this is what I would like to call the GIS-MRA techniques. Examples of recent studies were Curry et.al. (1990), Hamid & Ghazali (1991), Aalberts & Bible (1992), Roseirs & Theriault (1992), Hamid & Croft (1993), & Longley et.al (1994). With this small number of studies, we are able to fix some "bolts-and-nuts" of the techniques for providing a vehicle for various basic property valuation analyses.

The core of the GIS-MRA techniques is spatial-statistical analyses of property determinants: how the spatial effects of these determinants on property value can be tested empirically. The MRA component alone has two traditional functions - "correlation" and "prediction". The GIS component alone also has two. The first is its capability of providing an integrated database handling system simultaneously relating spatial and attribute data of properties. The second is its capability of spatial operation and representation of these data.

2.0 RATIONALES AND TENETS OF GIS-MRA TECHNIQUES

2.1.1 Why GIS-MRA Techniques

The use of statistics in conjunction with GIS has emerged out the needs for using aspatial analysis
by Haas (1922) & Wallace (1926). However, although the statistical milieu provides models that are capable of producing relatively accurate and consistent regression estimates, the technique does not possess spatial analysis capability. This capability is necessary because the incidences of property value are locationally distributed. Furthermore, many complex phenomena that change spatially and temporally can only be better understood by analyzing them in a visualized manner. GIS fills in this gap.

In the past, the traditional non-statistically based as well as statistically based valuation techniques did not treat spatial phenomena in an explicit way. While there is no compelling need to do so, the GIS-MRA techniques, when used, offer some advantages. The advantages are gained from using spatial statistics for real estate problem solving. Spatial statistics are used chiefly for spatial reasoning (Bartlett, 1975; Ripley, 1981; Clark & Haskins, 1986; Davis, 1986; Ansley, 1988; Odlund, 1988; Aria, 1989; Isaaks & Sivastava, 1989; Griffith & Amrhein, 1991). In the traditional MRA applications, valuers are making statistical reasoning through two traditional functions of the technique – “correlation” and “prediction”. The MRA has been applauded as scientific (Wallace, 1926; Renshaw, 1958; Davis, 1965; Shahan, 1973; Bruce & Simrell, 1977), having some essential scientific characteristics (Mundy, 1992). One of these essential characteristics is that, MRA technique identifies specific value factors and explicitly quantifies them. However, much of the data required for quantifying these factors are spatial data. Consequently, the two functions are performed without, as indicated before, an explicit reference to spatial dimension.

On the other hand, GIS has an enormous spatial functionality. The dominant feature of GIS is its capability in the numerical representation of spatial information, thus, providing a wealth of quantitative processing. Also spatial statistics are capable of characterizing the geographic distribution, or pattern, of mapped data. These statistics describe spatial variation in data, explicitly. Spatial statistics incorporate location information in displaying the variation of thematic values. The distribution is characterized in geographic space rather than numerical span, as in the traditional statistics (see Batty, 1987). Put it another way, the spatial statistics, that include spatial data analyses reduce spatial patterns to a few clear and useful summaries. There are so many different types of spatial patterns that we need to summarize the data in one or more graphs rather than by single numbers, such as the mean and standard deviation of classical statistics (Ripley, 1981: see p.1).

As GIS is considered to be a versatile system for a wide range of applications, valuers and appraisers are strikes with the idea to combine it with MRA for real estate decision making. Fundamental in the decision making is the analysis of complex phenomena that change spatially and temporally. In this context, valuers may like to answer, confirm, or rationalize certain questions or issues regarding the influences of some spatial elements on property values (see Section 3.3). With GIS, such questions or issues can be analyzed spatially. Then, spatial hypotheses can be tested statistically. With the proper framework, the GIS-MRA techniques can be used as a new tool of spatial reasoning in real estate analyses.

3.0 MAJOR STEPS IN GIS-MRA TECHNIQUES

There are many ways in which valuation analyses can get the service of statistical modeling and GIS functionality. These ways are by no means complete and, thus, should only be illustrative of the multitude of problems the GIS-MRA techniques can solve. The steps in applying GIS-MRA techniques include objectives and problem identification, the functional requirements of the techniques, the intended analyses, and other considerations.

3.1 Objective and Problem Identification

If one believes that the capability of GIS-MRA techniques can improve valuation analyses, for the
GIS-MRA Techniques in Property Valuation: A Framework for Implementation

procedures at hand, this can be the starting point towards their application. Then, defining the problem is a matter of recognizing the nature of the situation. In the professional practice, the problems to be solved and objectives to be achieved relate to the actual "term of reference" from the client or, what is considered as "improvement of services" to the public. In property research, the problems and objectives depend on "what" the researchers wish to add to the profession, no matter how small they might be. They may, at least, have some forms of academic contribution, but may also be of practical use.

The objectives can range from as general as establishing a land information system (Lockwood, 1989) to relatively specific ones such as for rating administration (Azhari & Mohammad, 1992), supporting property mapping, appraisal/valuation process, and/or property management systems (Li & Ruggiero, 1991, ARClNFO News, 1991), monitoring land movement and future trend forecasting of land price (Olmo, 1985) to even a much narrower ones such as the normal appraising with the emphasis of geo-spatial analysis technique (Weber, 1990) and spatial hypothesis testing of location effect for a trend surface analysis (Hendel & Ingham, 1981).

Since the above objectives can still be too general for a specific operational process, an appraisal or Valuation process may need to define the specific problems to be tackled. There is no hard-and-fast rule for identifying these problems. All, with the guide of the objectives, must be user-defined: "something" to be achieved from an appraisal/valuation exercise. Examples are given in Table 1.

<table>
<thead>
<tr>
<th>Objective</th>
<th>The use of GIS-MRA system for property assessment administration of a country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Identified</td>
<td>How</td>
</tr>
<tr>
<td>To show commercial vacant land sales of the country</td>
<td>Map location of vacant land sales, each with a circled red number. Look-up description of the sales.</td>
</tr>
<tr>
<td>To show industrial vacant land sales for each of two years</td>
<td>Map location of vacant land sales in contrasting colours. Ovals show the sales take place and the price range for various locations</td>
</tr>
<tr>
<td>To show the major places of concentration of customers and public</td>
<td>Map location of department stores by chain and all enclosed audits in colour coding.</td>
</tr>
<tr>
<td>To show and correlate the distribution of apartment income and location</td>
<td>Map location showing market income for apartments. Correlation between apartment income and location.</td>
</tr>
<tr>
<td>To identify market trend and estimate neighbourhood boundary between low and high income areas</td>
<td>Map showing distribution of low and high income area. &quot;Hot value&quot; map plotting annual average percent change in just value in these areas.</td>
</tr>
</tbody>
</table>

Table 1 Objective and problem identification in Appraisal Valuation Exercise (constructed from Ashby, 1992)

It must be noted that the objectives should not reflect only "what" to be achieved as the "end product" (such as a map) but also, "how" to deal with it (such as, data handling, capability). The latter issue is, in fact, one of two major reasons why the GIS is needed. This is the capability of providing an integrated handling system simultaneously, relating spatial and attribute data of properties. Although it carries the same message of "data marriage", the way it is implemented and the resources needed vary from one situation to another as reflected in a few studies (Fors & Atkin, 1989; Curry et al., 1990; Humel & Ghazali, 1991; Rosier & Thorp, 1992).

3.2 The Functional Requirements of GIS-MRA Techniques

From the preceding discussion, two principal categories of functional requirements of GIS-MRA technique can be identified. The first is the GIS functionality itself (see Thrall & Marks, 1993). This

65
functionality, with different degrees of utilities and sophistication, is contained in a particular GIS software. Thus, evaluating GIS software must be one of the major considerations in using these techniques (Marks et al., 1994). [Grumbles among property people saying, for example, "this (GIS) software is monstrous and learning (thousands of) its commands is a total waste of time" are sometimes heard]. The second is the statistical modeling functionality. There are many established statistical software packages such as the SAS, SPSS/PC, SHAZAM, STATA,TIBS, and GENSTAT that can conveniently serve modeling purposes. Three packages are external to GIS software and, thus, not subject to software's limitations.

The functional requirements, if properly identified, form the backbone structure of the GIS-MRA techniques. The references cited under Section 1.0 can be used to frame up this structure. First, it requires system resources. Second, it requires some data communication protocols. Third, it requires data integration. Fourth, it needs statistical modeling. Then, all of the above requirements must be tailored towards the defined or intended validation analyses.

3.2.1 The System Resources: Hardware-Software Set-up, Data and Personnel

To successfully deal with complex nature of geographic data, a GIS should have at least three major components that make a balanced system - computer hardware, sets of application software modules, and a proper organizational context (Burnough, 1987). Contemplating only on the first and second components, the very essence of a GIS comprises hardware and software, designed to manage spatial databases (Lee & Zhang, 1989).

The computer hardware generally comprises the following components and functions: central processing unit (CPU) - input, editing, GIS commands, etc; storage/memory unit - data and program storage, visual display unit (VDU) - user-operated computer and peripherals control, digitizing unit and for data conversion facilities and, plotter - presentation of data. The software package of a GIS consists of five basic technical modules: data input and verification, data storage and database management, data transformation (or manipulation and analysis), data output and presentation, and interaction with the user (Burnough, 1987). Within these modules, there are several operational components, which vary from one software to another. The ARC/INFO software, for example, consists of twelve major components: Arc, Info, Arcedit, Arcplot, Librarian, Tables, Network, Route, Allocate, Grid, TIN, and COGO.

Data input converts data from their existing forms into one that can be used by the GIS; data management, includes: functions needed to store and retrieve data from the database, data manipulation and analyses determine the information to be generated by GIS; and data output can be in the forms of maps, tables of value, hard-copy texts such as papers, or soft-copy texts such as electronic files (Aronoff, 1991 : see pp. 42-43). These basic components ensure the efficiency of the overall system and provide a fully integrated database that will support administrative and decision making functions.

Valuable data consist of spatial and non-spatial data components. In a broad definition, these two types of data can simply be referred to as geo-referenced data or geographic data. Geographical data are most commonly thought of as having two basic characteristics: (a) the actual phenomenon such as the variable, its classification, value, name, etc.; and (b) their spatial locations i.e., the locations within geographic space where they reside. Another important characteristic which is particularly relevant to GIS is that of time (Dangendorf, 1990). According to him, geographic data may or may not change over time. Since property market is dynamic and, thus, likely to change, the spatial and/or temporal changes in the data representing certain market events can be conveniently depicted using GIS.
In terms of spatial characteristics, geographic data are complicated; they must be described in relation to locations or positions, topological connections, and attributes of the objects recorded. Geographic data are referenced to location by using standard system of coordinates, may be local, national, or international. Geographic phenomena are represented by points, lines, or polygons (areas) plus labels. Labels could be the "actual names, numbers cross-referenced with legends, or symbols" (Burrough, 1987: p. 17). Spatial data represent the geographic location of features (e.g., roads, parks, etc.); while the in-spatial/attribute data provide descriptive information (e.g., street names, riv names, names of place, etc.). During data inputs, the spatial and attribute data must be entered and correctly linked, i.e., the attributes must be logically attached to the feature they describe.

Obviously, having defined the objectives, values must examine the hardware-software set-up and sita, in specifying resource requirements for their analyses. This is necessary to avoid frustration or failure. Data issue can be one major motivation as well as discouragement to using GIS-MRA techniques. It is quite tempting, because of the enormous integrative capability of GIS, to amass a wide range of various data from the existing sources (digital/analog). Both have some advantages and disadvantages. Manual digitizing is cheaper but time-consuming and error-prone2, while "ready-made" digital data are of better quality but, often, can be too expensive3. Thus, the decision on data can be largely influenced by the availability of funds. The hardware-software system requires a high initial capital outlay. Even if this is resolved, the individual GIS users will still be faced with side issues such as system familiarity (e.g., PC-based versus network-based systems, simple versus sophisticated GIS packages), professional advice, in-house training or continuing professional development, and the existence of peer support group.

GIS-MRA techniques may be based on an open approach, which has two main benefits (Rosiers & Theriault, 1992). First, the data transfer philosophy combines the power and functions of many individual low-cost packages on desktop computers to attain the performance and flexibility of large-scale workstations' mainframe systems. Second, the system architecture is highly modular. As such, we may add new functions when needed and replace some packages by other packages. More important, it is also possible to transfer data to many other computers and development platforms. Despite the relative complexity of data transfer, this approach is very powerful because we can integrate new functions as they become available and we do not rely on specific manufacturer. Even more valuable is the ability to integrate cartographic databases, digitized by various public departments in many forms and systems, although they are not always compatible.

3.2.2 Data Communication Protocol and Analysis Platform

The fact that "GIS, have grown out of a number of other technologies and a variety of application fields" (Martin, 1991: p. 28) has led to an important phenomenon. That is, there is no single GIS which could exclusively fulfill the analysis requirements of diverse applications. For example, the mass valuation, where statistical modeling is involved value-estimating equations need to be derived externally using some specialized statistical software as mentioned earlier. Even the basic property data are stored in different systems and data formats. Besides, some basic geographic data required for valuation analyses are also stored in different spatial formats.

Various hardware-software systems were originally developed to serve different purposes. This situation necessitates efforts to bring together the relevant hardware-software-data systems, to work in unison, for the intended valuation analyses. This is where a data communication protocol is fundamental. Since most data links in GIS-based valuation analyses adopt an "open" concept, there are essential features to be identified, to ensure a generic solution to creating GIS-MRA models. These features are (Curt, et al., 1990):

• The CAMA system must be capable of creating new data fields or adapting "open" fields for use by incoming data from the GIS. These data elements could be used in later modeling efforts.
Ayd. Hamid bin Mar Iman

- The CAMA system must be able to produce an ASCII file from data stored in its files.

- In order to store incoming data rapidly in the appropriate records, the CAMA system must be able to create and use an index built on parcel identifiers. The users must have built such index before data are imported.

- Either a GIS must have a user programming language enabling users to access an incoming ASCII file and attach data from it to specific polygons, or it must have a module expressly designed to read such a file (with key parameters set by users) and place the data appropriately.

- The GIS must be able to associate key descriptors with its polygons and, in turn, associate other data elements with those descriptors.

- The GIS must be able to create new, synthetic variables derived from its analysis, associate these with particular polygons, and then translate them into ASCII format transmissions to the CAMA system.

- The link should allow any CAMA programs to exchange data with any GISs. This requires the programs to have sufficient flexibility so that common data elements can be defined that allow one program to recognize and associate data with the entities of the others.

- Data should be transferred by means of ASCII files rather than by one program reading the internal files of the other. This allows complete flexibility in the choice of GIS or CAMA system.

- A common naming convention should be used for the ASCII transfer files.

- Sequential files are preferred, because they will automatically be rewritten each time they are used. However, CAMA and GIS programs must include information on the number, types, and names fields (or factors) being transferred in the file stream.

- Once data from either system are moved into the other, there should be no a priori assumptions in the program about what data elements (beyond the parcel ID) will be transferred or where they will be placed. It should be assumed that every data element received from one system will be used in the other.

Given the data sets, GIS-MRA analyses rely on the interaction of several software packages and hardware. Thus, any GIS-MRA valuation analyses have to identify the suitable packages and data transfer approach to transfer information from one package to another. This points to the requirement of an analysis platform. Based on Johnston (1992), there are three platforms of GIS-MRA analyses. First, extract the spatial GIS data, pass them into a statistical package, and transfer them from the statistical package to the GIS. Second, include either all statistical functions within a GIS or all GIS functions within a statistical package. Third, include necessary or complementary statistical tools directly into GIS software. With mutual exclusiveness of statistical tools and GIS software, it is much easier to adapt the first analysis platform.

3.2.3 Data Integration

The chief aim of valuation data integration is to enable concurrent data analyses in a simple environment such as the ARC/INFO (Hamid & Harrington, 1994). A big mass of data required for value estimation can only be analyzed concurrently by an efficient technique that is capable of
converting and configuring data from a myriad of sources and types of systems into the required formats for an informed valuation analysis. In the automated systems, data convertibility and configurability directly determine the success of data integration. Being maintained in different systems, these two aspects are expected to vary in degrees. Upon successful integration, GIS is able to function in supporting concurrent analyses of diverse data of different forms, sources, and storage, pertinent to valuation. These data may include such as those from valuation rolls, cadastral records, soil maps, topographic maps, geological maps, climatic maps, and socioeconomic maps (population, income, utility, etc.). The general procedure for integration is quite similar to that for the agricultural suitability assessment (see Laar, 1992). The integrative capability of GIS has made it a major tool for working through the pool of information that could become an ocean of information (Francis, 1991). Thus, being achieved as a single analysis, gives a GIS a synergistic power – the increased capability realized when components work in unison (Vale & Bennet, 1991).

In New Zealand, the technique called "parsing-and-matching" is the bridge of success for integrating the two principal databases the Digital Cadastral Data Base (DCDB) and the VALPAK’s sales information. Crott et al. (1992), Hamid & Harrington (1994) have presented the relevant examples of the integration in the mass valuation of farm properties. Hamid & Harrington (1994) remarked that integrating sales and cadastral database for valuation purposes is like engaging a couple prior to a marriage. It is the very first move in bringing two parties into a union. Again, in the New Zealand’s case, one major obstacle towards achieving a successful data integration is the non-uniformity of the systems of parcel identification adopted by both the Department of Survey and Land Information (DOSLI) and Valuation New Zealand (VNZ). Assuming that the current DCDB system adopted by DOSLI and the VALPAK system adopted VNZ remain the same, a greater rate of success of data integration may be achievable only by the parsing-and-matching algorithm. The manual searching-and-matching is still needed but, perhaps, less painstakingly.

3.2.4 Statistical Modeling

Valuation analyses can be statistically or non-statistically based. Each of these choices has some requirements. Statistically based analyses demand modeling step. Statistical modeling, in itself, is already quite an undertaking. An MRA model can be employed exclusive of the GIS. However, when the combination is desired, the statistical modeling is a more crucial part. The reason is, in the context of the GIS-MRA techniques, a poor model can be a cause of an over or under caused by poor spatial analyses. The first key words imply a situation where by a poor MRA model is used to explain to predict spatial phenomena (e.g., to predict the market values of unsold properties). The second key words imply a situation whereby poor spatial analyses result in a poor MRA model. For instance, because of a slavishly analysis of the spatial pattern of farm value along a railway and/or a highway, a modeler may decide not to include dummy variables representing the effects of those elements, in his model, while, in fact, these effects are significant.

A decision of primary importance in the modeling process is that of choosing variables to include in the regression equation (Mark & Goldberg, 1988; Pace, 1991). It must start with the analyses of value factors. The process concentrates on selecting the most pertinent variables in general, the economic theory often provides detailed guidance on the dependent and independent variables that should enter a model (Pare, 1991). In practice, however, the actual sets of variables included in the models are different from one case to another. The final choice of variable selection is, thus, still problematic. Tentative inclusion of variables into a model can be made based on the following steps:

• Consider the literature at hand, no matter how inconclusive it might be, and make one's own decision. Using the information, added with one's personal knowledge and/or experience about the local situation, rank the most pertinent value factors.
• Confirm the list with any previous published information relevant to the locality and type of property, or by asking opinions of other counterparts.

• Divide the variables into certain categories, for example, "focus", "free", and "doubtful" variables (Graves, et al., 1988). Focus variables are those of particular policy interest, although they may in fact prove to be doubtful (e.g., zoning, planning permission, flood, land quality). Free variables are known to affect the dependent variable (property value) but are not specific interest (e.g., weed, disease, buyer/seller equations). Doubtful variables may or may not affect the dependent variable, and the direction of the effect may be uncertain (e.g., local crime rate, income level of the purchaser).

• From the integrated databases, conduct careful spatial analyses. This step is useful in many ways. It can be used to give a priori information on some hypothesized value factors. Apart from it, GIS helps performing basic spatial functionality (see Thrill & Marks, 1993). For instance, it can be used to reclassify neighbourhood boundaries or parcels. It can also be used to identify and or calculate certain (possibly new) variables such as the distance of parcels from population centres, land slopes, existence of floodplain, power lines, etc. All these help model specification.

• Include the candidate variables in the regression equation and perform simple correlation analysis. For this purpose, a subjective multicollinearity cutoff value of correlation coefficients can be set at ±0.8000 (King & Sinden, 1988). If a value is found to be greater than ±0.8000 it is still acceptable. A value equal to or greater than ±0.8000 can be said to reflect a serious multicollinearity issue.

• Test run the model, using a pilot sample, and evaluate the results. The purpose is to ascribe whether a particular variable is part of value factors, whether it makes sense, and whether it is a significant value contributor. Where possible, go back to the spatial analyses for cross-check purposes. Combining the initial statistical results and spatial analyses may help improve modeling (e.g., it may lead to model re-specification).

Care needs to be taken on two aspects of variable inclusion. First, include only the pertinent factors. The regression test run may be able to help sort out these factors. Second, not only factors whose influences could be assigned to each and every land parcel (at micro level) should be specified. Factors such as the consumer price index, interest rate, land price index, and building price index that exert effects on a macro basis (on aggregate farms), also need to be included. The reason for the first one is that, while the addition of minor and possible redundant variables may increase the coefficient of determination (R²) only marginally, such an increase may not justify the burden of additional data collection and reduction in the degree of freedom. As for the second one, the inclusion of macro regressors is to evaluate the importance of macro effects on property values. Many macro regressors are economic variables that are normally influential in the property market. In most valuation cases, at least, one of these macro factors such as the inflation rate, mortgage rate, and consumer price index, exerts significant effect on farm value.

The next consideration is the correct choice of model. This has been asserted in many studies (Colwell, et al., 1983; Reichert & King, 1986; Couison & Robins, 1987; Mark & Goldberg, 1988; Murphy, 1989; Brozman, 1990; Weirick & Ingratz, 1990). These studies implicated the need for hypothesizing and examining various functional forms of property value models. This is achievable introspectively and retrospectively. That is, valuers analyze the expected regression model by some logical reasoning, guided by previous studies in the similar problem areas.
Still within the consideration of functional form is the question of single-equation versus multi-equation models. Many MRA models in valuation studies were specified as single-equations (Haus, 1922; Wallace, 1926; Ezekiel, 1936; Ahmed & Parcker, 1964a, 1964b; Davis, 1965; Abdel-Badie & Parcker, 1967; Hilderbrandt, 1968; Pen, et al., 1968; Crowley, 1974; Wise & Dover, 1974; Bell, 1977; Clifton & Sparrow, 1983; Woolford, 1983; Gibson, 1988; Payne & Tindell, 1993; Mitchell, 1994). Only a few studies employed multi-equation models (Tweiten & Martin, 1966; Reynolds & Timmons, 1969; Hamid, 1993; & King & Sidden, 1988, 1994). In each category of the equation forms above, there are various MRA variants. This shows that, choosing models to use is never a simple task.

The major outcome expected of a modeling exercise is a reliable and plausible model with a good predictive capability. In relation to the second aspect, it can be said that “the science of model-based) valuation lies in formulating and implementing models with high predictive accuracy” (Lueke & Horsley, 1991: p.28). Therefore, real estate modeling is concerned with the estimation of parameters within some specified models that can meet this expectation. A variety of parameter estimation techniques are available. The most common one is the Ordinary Least Square (OLS) technique.

The final step is for a property value model to be diagnosed against the assumptions of OLS. The most crucial are multicollinearity, heteroscedasticity, autocorrelation, and mis specification. Upon passing the battery of these tests, the model is ready for further applications.

3.3 The Intended Analyses
3.3.1 Some Previous Analyses
Some previous works in natural resources (Pereira & Hami, 1991; Johnston, 1992) and real estates (Curry, et al., 1990; Hamid & Ghazali, 1991; Rosiers & Theriault, 1992; Amari & Ghazali, 1993; Hamid & Croft, 1993) have shown some examples of spatial analyses, incorporating GIS and/or MRA.

In Hamid & Ghazali’s (1991) study, the MRA model was used to create a value map of sold and unsold properties. This is particularly useful for valuation analyses especially in evaluating value profile across a geographic area. Besides, as they noted, the map is useful for some decision making regarding land resource management.

In Rosiers & Theriault’s (1992) study, putting aside the impressive potential with respect to data management, a major and particularly promising feature of GIS-MRA technique was its ability to provide an analytical framework for information of a thematic aspatial nature (e.g. property prices, average household income, and effective property tax rates), which can be superimposed on a geographic support, providing some link between the former and the latter. This has allowed spatial cross-analyses of various phenomena, which added substantially to the already great potential of the statistical analysis tool.

In Aaheber & Bibe’s (1992) study, the hypotheses tested were connected to the question of whether race, income, and distance from the central business district (CBD) were related to the incidence of default. Correlation and regression analyses were applied using census tracts as observations. The dependent variable was the percentage of defaults relative to the total number of houses per census tract. The independent variables were the percentages of persons who were African-American within each of the census tracts as well as family income. Distance in miles from the CBD was also tested to discover whether defaults could have decreased with distance from the suburbs. Besides, several other census-related socioeconomic variables were also tested. The results, however, showed that none of variables have supported the hypotheses.
Although statistically significant results did not emerge in their study, the authors pointed out some advantages of using GIS in the context of a typical real estate database. First, with an extensive mapping capability, a major advantage of the ATLAS GIS was its capability to merge databases with geographic information or spatial analysis with the resulting mapping and data information. Second, many types of geographic analysis can be made possible, including the grouping and clustering of data according to quality or quantity characteristics (e.g., size and sale price) found in the database, or by geographic characteristics such as distance from a point on the map or distance from a feature such as floodplain. Third, the selection and condition functions were extremely useful for observing data that conforms to certain characteristics, such as home values in a particular value range. Fourth, the thematic mapping functions were readily applicable to mapping by area, especially census data by census tracts, other geographic areas such as zip codes, or areas specified and custom-down into the geographic map by the user.

In Curry et al.'s (1990) study, the effects of intrinsic property characteristics (lot size, total living area, number of rooms, number of bathrooms, garage space, age, physical depreciation, roof type, heat air conditioning, and holding age) on property value were exclusively focused. Few other extrinsic variables were also considered. They included location (in or outside town), distance from a defined point in town, population of town, and the existence of bank. Their study showed that the use of GIS has enhanced the understanding of the spatial patterns of the effects of those factors on property values.

3.3.2 Spatial Hypotheses and Model Specification

For the purpose of our discussion here, a spatial element is simply defined as an entity that can be geo-referenced on a map, such as a land parcel, a town, a railway or highway. In the GIS jargon, as indicated under para 3.2.1, these are examples of spatial objects which are characterized by x-y co-ordinates, point, and line, respectively, plus their labels.

Again, a variety of spatial questions and hypotheses can be posed in the process of explaining the spatial "behaviour" of property value. In essence, GIS can visually analyze the pattern of the behaviour with regard to a very important factor, i.e., location! location! location! (Pearson, 1991). Since spatial objects such as population centres, railway or highway lines are locationally fixed, values must analyze them in respect of their given locations. From here, spatial hypotheses can help in the statistical modeling, especially with regard to variable inclusion in an MRA model. The kinds of spatial analyses used for testing the hypotheses include proximity, buffer analysis, and neighbour re-classification.

GIS-MRA techniques give the opportunity for the statistical and spatial analysis components to work in a two-way direction. First, the statistical model utilizes "a priori" spatial information from GIS maps. Then, upon generating statistical results, the information can be subject to revision for model re-specification.

3.3.3 Spatial Prediction of Property Value

The GIS-MRA techniques form a basis for spatially based property value depiction. Let I denotes data themes of layers defined by the MRA variables, P property parcel, and X the numeric value of property characteristics associated with each parcel. The layer multiplier B, is derived from the regression equation. Given the spatial (mapping) set, each layer will consist of an array of property value's component sub-products as follows:

72
GIS-MRA: Techniques in Property Valuation: A Framework for Implementation

Data layer Property parcel

\[ P_1 \quad P_2 \quad P_3 \quad \ldots \quad P_n \]

\[ \beta_1X_{11} \quad \beta_2X_{12} \quad \beta_3X_{13} \quad \ldots \quad \beta_nX_{1n} \]

\[ \beta_1X_{21} \quad \beta_2X_{22} \quad \beta_3X_{23} \quad \ldots \quad \beta_nX_{2n} \]

\[ \beta_1X_{31} \quad \beta_2X_{32} \quad \beta_3X_{33} \quad \ldots \quad \beta_nX_{3n} \]

\[ \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \]

\[ \beta_1X_{m1} \quad \beta_2X_{m2} \quad \beta_3X_{m3} \quad \ldots \quad \beta_nX_{mn} \]

Note that the "\( \ldots \)" separating each parcel indicates that each and every parcel is unique and independent of the others, so that in the data analysis, no inter layer mathematical operations can be performed within the \((n \times m)\) parcel-layer. For example, in manipulating data for layers, the data values for the \(n^{th}\) Parcel cannot be multiplied with, subtracted from, or divided by those of the \((n-1)^{th}\) parcel. So, the above data array is, strictly, not an algebraic matrix.

For each particular layer, the numeric values of property values’ component sub-products associated with each parcel is derived by multiplying \( \beta's \) with \(X's\) of the parcel. For example, for parcel \( P_1 \):

\[ P_1 \cdot \beta_1X_{11} \]

\[ P_1 \cdot \beta_1X_{12} \]

\[ P_1 \cdot \beta_1X_{13} \]

\[ \vdots \]

\[ P_1 \cdot \beta_1X_{1n} \]

So that,

\[ L_1P_1 = \beta_1X_{11} + \beta_1X_{12} + \beta_1X_{13} + \ldots + \beta_1X_{1n} \]

\[ \quad = \sum_{x=1}^{n} \beta_1X_{x1} \]

Such an algebraic operation applies similarly to other individual parcels in the mapping set. More generally, for a particular parcel \( J \), its corresponding numerical value on the final layer, \( L \), which is an algebraic summation of \( n \) sub-layers is given by:

\[ L \cdot P_j = \sum_{x=1}^{n} \beta_{xj} \text{ for } i \neq j \]

Note that \( L \cdot P_j \) = market value of property \( j \). In words, the numeric value of the final layer, \( L \), is the model-estimated market value of property parcel \( j \). So, basically, the parameter estimates in the last equation will be input to GIS software for generating value maps of farm properties or any other spatial depictions.

3.3.4 Spatial Analysis of Residuals

One aim of this analysis is neighborhood re-classification. Quite often that property sub-markets surpass the physical boundaries initially used for delineating neighbourhoods. Since the (possibly new) boundaries of these sub-markets could not be identified, property value profile \( \omega \) be used to help identifying the possible re-classification of neighbourhoods. Two ways are possible using GIS-MRA techniques.
The first way is to form contour map of spatially interpolated model's predicted residuals. This map is to be produced by way of computing distance effect on a triangular irregular points network to produce, say, a 200-meter-square regular grid. The resulting map is then plotted using, say, $20000$-contour intervals. The contours can be used to spatially examine areas exhibiting "steep" surface of value residuals. These are areas over which the MRA model is expected to be less accurate for prediction, because property value changes drastically from one point on the residual-value gradient to another. Such an analysis enables valuers to further investigate the likely elements that hinder more accurate prediction of property value within and across neighbourhoods such as spatial autocorrelation.

The second way is to follow the manner in which Sooh and Schütz-E (1966) classified neighbourhoods of farm properties. According to them, the computed regression residuals quite frequently show systematic patterns that suggest the addition of another variable. On the basis of similarities in the sign of residuals the parcels can be arbitrarily divided into a number of neighbourhoods. The dummy variables representing the new re-classification of neighbourhoods are then entered in the value estimating model and regression re-run is performed to see whether the results can be improved.

3.4 Other Considerations

Although cartographical support is an "outdoor" aspect of GIS-MRA techniques, it can adversely affect implementation, especially in terms of funding and data acquisition. The host institution may not be able to support data purchase. As mentioned earlier, digital data can be too expensive and may not justify the end results. The agencies holding/possessing the essential data may be reluctant in many respects. Some government agencies insist on "status-quo" of their data management systems. Because these systems are not fully compatible with each other, problems such as data exchanges and integration can slow down problem-solving process. Lack of communications and understanding may halt innovative efforts in the application of the techniques. The practical use of the technique needs to be put on trial. Since the GIS-MRA techniques can be costly and slow in adoption, a long-term plan may be needed to justify their use. Cost-benefit aspects needs to be studied in greater details. All these are other considerations in the evaluation of the techniques.

4.0 CONCLUSION

This paper has discussed a simple framework on how GIS-MRA techniques can be implemented for real estate analyses, in particular, property valuation. The core of the techniques is the combined use of GIS and statistical models for improving spatial analyses of real estate phenomena such as property values. Nonetheless, other fundamental element of the technique is the capability of the GIS component in providing an integrated database handling system, simultaneously relating spatial and attribute data of properties in a single analysis environment.

The GIS-MRA techniques, if carefully employed will be able to forge a new scientific dimension of real estate analyses. They are a tool that valuers cannot simply overlook.

ENDNOTES


2 The forms in which geo-referenced data, especially spatial data, are arranged and stored in computer is known as data structure. These forms encompass files and data access and database structure. The common forms of GIS data are simple list, ordered sequential files,
GIS/ARIA Techniques in Property Valuation: A Framework for Implementation

and indexed files, while those of database structure are hierarchical, network, and relational. Since GIS-based valuation analyses are "the other side of a coin" of GIS (i.e., an application side), these themes, which are the methodological side of GIS theory, are not an important aspect for discussion in this paper. Special references on these themes can be found for example, in Burrough (1987); Gerring (1984); Cook (1982); Drey (1981); Bootter (1980); Haralick (1980); Ullman (1980); Cook (1978, 1983); Walker et al., (1986).

I estimated that a 10,000-to-25,000-parcel manual digitizing task will take approximately six month man-days with a 5-10m positional accuracy level, relative to the topographic survey coordinate multiplied by the representative fraction of the capture map scale (see DOSU, 1994: pp. 2-7, 2-10).

Let's take a New Zealand's example. The ex-GST rate for the Digital Cadastre Data base (DCDB) is NZ$3,500 per parcel. For a study covering a reasonably large area comprising 15,000 - 15,000 parcels), the digital cadastral data alone would cost NZ$3,000 - 4,500. Other GIS data needed for reasonably comprehensive analyses are Digital Mapbase (MGB) with population/with statistics (NZ$3,000 - 4,500); New Zealand Resource Inventory (NZRI) free-of-charge; Digital Topographic Data; DTD such as contours and spot heights only) NZ $4,000 - 6,000). Thus, depending on the details and extent of the data required, a ten-to-fifteen-thousand-parcel study area may cost NZ $10,000 - 25,000, ex-GST, materials, translation, mailing costs, etc.

A poor model is a model which is buried by statistical problems: multicollinearity, autocorrelation, misspecification, and other violations of statistical assumptions. For a cogent discussion, see Cook (1988). It must be qualified that, even without GIS, a poor MRA can still result or even without GIS, a poor model analysis may still crop up. Therefore, the statement must be interpreted strictly within the context mentioned in this paragraph.

Most GIS textbooks contain discussions on these concepts. Consult, for examples, Martin (1991) and Antruff (1991). Walker, et al. (1986) use a more general term "feature" to refer to a fundamental geographic entity in the database, which has the following characteristics: (1) one or more geographic coordinates; (2) an optional associated text string; (3) optional graphic parameters; (4) symbolic attributes; (5) optional user-defined attributes.

REFERENCES
ARCHINFO News (1991) "DIAL and Sim Commercial Bank Develop Property Valuation and Management System", Redline, California. USA publication, Fall issue.


Pen, J.D. et al. (1998) "Farming and Market in the Mississippi River Delta Cotton Region", Texas A&M: Louisiana State University, Department of Agricultural Economics Report 372.


Reynolds, J.L. & Thoma, D.J. (1969) "Factors Affecting Farmed Values in the United States", Iowa State University Agricultural Experiment Station Bulletin No. 556.


GIS-BRA Techniques in Property Valuation: A Framework for Implementation

Scheib, G.J. & Schnack, V.C. (1966) "Quantitative Analysis of Some Farm and Non-Farm Determinants of Agricultural Land Values - Impact of Economic Development", Purdue University Agricultural Experimental Station, Research Bulletin No. 821, Lafayette, Indiana, November.


