A comparison of land valuation methods supported by GIS

Nair Cristina Margarido Brondino
Antônio Nélson Rodrigues da Silva

UNIVERSITY OF SÃO PAULO
São Carlos School of Engineering - Department of Transportation
Av. Dr. Carlos Botelho, 1465
Phone +55 (16) 273 9595 - Phone/Fax +55 (16) 273 9602
13560-250 São Carlos - SP - BRAZIL
E-mail address: anelson@sc.usp.br

ABSTRACT

The purpose of this work was to study three different strategies for the appraisal of urban land. The first, a theoretical strategy created by the authors of this study to reproduce the common conditions of Brazilian cities, uses increments and reductions in the value of a square meter of land according to each lot's individual features. The second method, based on Multiple Regression techniques, is widely used for valuation purposes. Finally, the effectiveness of Artificial Neural Networks to deal with this kind of problem is studied. A sample of 157 lots was collected from several neighbourhoods of a small Brazilian city for the case study. The lot features recorded were area, width, shape, distance to the downtown district of the city through the street network, existence of fences and paved sidewalks, and market price. Prediction errors have been estimated for each of the three methods in order to compare their results. Predicted and error values, added to Geographical Information Systems, may be used to build thematic maps and to check how each strategy applies to different areas of the city. The analyses of error values conducted in this study showed that Artificial Neural Networks presented the best performance as a land appraisal method for the case studied.

1 INTRODUCTION

The estimation of urban land values for taxation purposes is a problem that has been broached in several studies and discussions in the international literature for several years (Bentick 1996; Lim 1992). Several authors have tried to build mathematical models to help local governments make fair calculations of land tax values. These models have varied from very simple approaches to highly sophisticated strategies, but they have all had a common goal, that is, they have always attempted to minimize the differences between estimated and real market values. Indeed, this is the first step to ensure a fair land taxation system.

Some of the simplest models currently used to estimate urban land values make use only of weights, which depend on the property's characteristics. An international example of this kind of model is the work of Azar et al. (1994), in Beirut. This same approach is employed in many Brazilian cities, such as Araçariguama (Gazeta de Araçariguama 1995), which is the case studied in this paper.

Another form of land appraisal that is widely used is based on multiple regression models. This method has been successfully used by researchers in several countries, for instance: Gibb, 1994, in Scotland; Aoki et al., 1994, in Japan; Pasha and Butt, 1996, in Pakistan; Abelson, 1997, in Australia; Phang and Wong, 1997, in Singapore; and González and Formoso, 1995; Vertelo, 1996; and Raia Jr. et al., 1996, in Brazil. Although the multiple regression technique can provide excellent estimations, the search to identify the best model requires a knowledgeable and experienced person. Furthermore, if the analyst ignores the relationships between independent variables, this can jeopardize the results, as well as problems arising from an interrupted flow of data. All these minor problems suggest that another method, devoid of so many restrictions, would be welcome.

Because of the shortcomings of traditional methods, a third technique emerged as a possible alternative to the problem of land appraisal: the Artificial Neural Networks (ANNs). Although the use of ANNs in this kind of application was not common practice until quite recently, a few examples showed some improvement in property value estimates when compared to conventional methods (Guedes 1994; Almond et al. 1997). Moreover, according to Falas (1995), Artificial Neural Networks were generally superior when compared to conventional statistical methods used for the same purpose.

An example of the good performance of ANNs can be observed in the work of Subramarian et al. (1993), who compared the technique with some statistical methods used for classification. They concluded that ANNs presented the best solutions in every circumstance, including small sample sizes and complex functions. Similar results can be found in Kwon et al. (1995), who used multiple regression models and ANNs to solve the traveler salesman problem. According to Tubb (1993), the main advantages of ANNs can be summarized as follows:

1. **The ability to learn from examples:** neural computers are able to learn from experience, improving their performance and adjusting to new and dynamic environments, as opposed to common computers;
2. **Reliability:** ANNs cope well with incomplete and fuzzy data, and they can also deal with situations previously unspecified or not encountered. They accept input flaws without immediately jeopardizing the accuracy of the next predictions. This behavior is different from the one found in conventional computers, in which the failure of one component means the failure of the entire system;
3. **High speed processing:** because Artificial Neural Networks consist of a large number of massively interconnected processing units operating simultaneously, they can process information faster than conventional computing systems;
4. **Software availability:** ANNs software are cheap and easy to find (it is even possible to find free software in the World Wide Web).

Despite all these advantages, however, the performance of ANNs for the purpose of land valuation has so far not received careful consideration in the literature mentioned earlier when compared to traditional techniques. It is, therefore, the starting point of this work, which studies the effectiveness of three Geographic Information

System (GIS)-supported land valuation methods. The main objective of this work is to confirm the good performance of new tools, such as ANNs, for this kind of application also. The first method is a variation of the very simple method used in the studied city. The method calculates the value of the land by a simple equation that allows for expansions or reductions in the original, standard land value according to variations in the parcel’s characteristics. The only change introduced here is the number of characteristics considered as value modifiers. The second land valuation method studied here uses multiple regression techniques. Finally, the effectiveness of ANNs is tested in this kind of application.

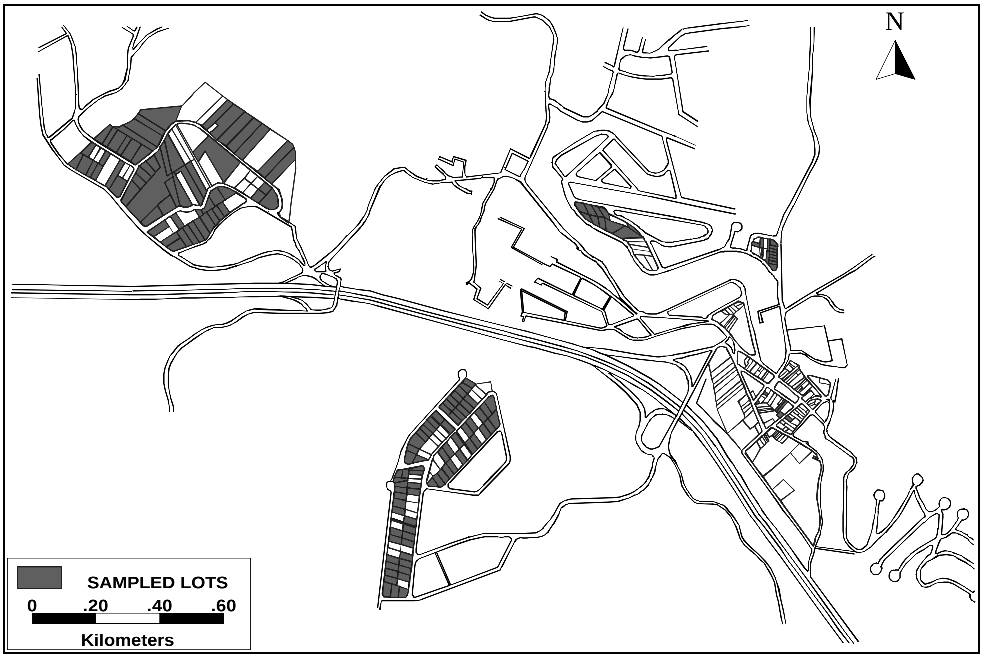
The city studied in this paper is a typical small residential town in the surroundings of the city of São Paulo, Brazil. Its main characteristic is the fact that most of its 6000 inhabitants work in the larger cities of the region rather than in the town itself. The total number of urban parcels is close to 1500 and includes both the built and the empty lots, of which 157 have been surveyed for this study.

This paper is divided into six parts. The section following the introduction contains a brief description of the methodology applied for data collection and analysis. Each one of the three following sections consists of a description and the results of the methods studied in this paper, i.e. the simplified land valuation method, the multiple regression approach and the Artificial Neural Networks technique. The results of the applications are reviewed and discussed in the conclusions of the last section, prior to the list of references.

2 DATA COLLECTION AND PREPARATION

The first step was to convert the street map of Araçariguama that until then had existed only on paper maps, into digital form and to store the data in TransCAD files. TransCAD is a Geographic Information System especially developed for transportation planning applications, although it is also suitable for the application intended here. A layer with the sampled land parcels has been created and real estate characteristics, such as area, width, shape, position in the block, distance to the city’s downtown area through the street network, existence of fences and paved sidewalks, and market prices have been input into its associated database. Some of these data have been stored as binary variables (1 when it is present or 0 if it is absent). The digital map of Araçariguama showing all the streets and sampled lots is shown in figure 1.

Figure 1: **Map of sampled lots and streets of Araçariguama.**



With all the data now in digital format, all the basic conditions for studying the three appraisal methods have been achieved except one: the original set of data has to be divided into three groups for use in the Artificial Neural Networks (ANNs) technique. To meet these requirements, fifty percent of the data has been randomly taken for the training phase, while two groups of twenty-five percent have been used for the validation and testing steps. This division has been done randomly three times and has generated three different sets of data. The relative estimation errors have been calculated for each one of these groups and their average is taken here as the total relative error. The same sets of data used to estimate the ANNs prediction errors have been used to test the performance of the other two techniques.

3 MODEL 1: WEIGHING BY PROPERTY ATTRIBUTES

This is a very simple model created by Brondino and Silva (1997) to reproduce and improve the process originally used in a Brazilian city. The original method was essentially the same, but it was based on only two variables: ground area and length of the lot. By the method adopted here, calculation of the property value is given as a function of an *X* variable that is equivalent to a square meter in a standard lot. The standard lot is defined as follows: flat, regularly shaped, located in the middle of the block, fenced, facing a paved sidewalk, located less than 500 meters from downtown, with a total area of 250 square meters and a width of twenty meters. All these references are modal values taken from the database containing lot features. In this method, characteristics that increase or decrease the property value are given as additions or reductions (in percentile terms) to the standard *X* value. The combination of all lot characteristics gives a correction factor. This factor, multiplied by *X* and by the ground area of the lot, gives an estimation of a property's value:

PROPERTY VALUE = PROPERTY AREA x *X* x CORRECTION FACTOR

For example, an *X* value found to be equal to 0.50, after observing all the characteristics of the property, means that the value of a square meter in that location is equal to fifty percent of the value of a square meter of the standard lot. The main characteristics of urban lots and the respective weights used here were obtained from real applications or from theoretical studies conducted by Brazilian researchers (such as Borges 1975, and Möller 1995). Among the characteristics considered important in calculating the value of land, those applied in this study are shown in tables 1 and 2, together with the weights employed.

All the calculations of this method were carried out directly with GIS built-in routines. Even the weights were introduced into the GIS database through logical sentences of the *if-then* type which were already available in the software. As a result, the total error obtained was 60.48. Considering that these are relative errors, this is a

significant advantages in terms of accuracy, although it is quite a simple alternative. It is easy to calculate, does not require very sophisticated procedures, and is also very

Table 1: **Weights of the variables of lot width, area, and distance from downtown.**

CHARACTERISTICS	**WEIGHTS**	
	Distance <= 1500 m	**Distance > 1500 m**
Area (m^2) + Distance from downtown (m)		
Area <= 1000	1.00	0.90
1000 < Area <= 5000	1.20	0.95
5000 < Area <= 10000	1.50	1.00
Area >10000	1.80	1.05
Width (m) + Distance from downtown (m)		
Width <= 20	1.00	0.70
20 < Width <=50	1.10	0.80
Width > 50	1.30	0.90

Table 2: **Weights of the variables of shape, location, area, and distance from downtown.**

CHARACTERISTICS	WEIGHTS
Shape	
Regular	1.00
Irregular	0.90
Location	
On the corner	1.10
In the middle of the block	1.00
Other	0.90
Improvements	
Fence or paved sidewalk	0,95
Fence and paved sidewalk	1,00
No improvements	0,90
Distance to the downtown area (m) - DC	
DC <= 500	1.00
500 < DC <= 1000	0.95
1000 < DC <= 1500	0.85
l500 < DC <= 2000	0.75
2000 < DC <= 2500	0.65
DC >2500	0.55

4 MODEL 2: MULTIPLE REGRESSION

The second method studied here is a model that uses multiple regression analysis (Draper and Smith 1981), which is the most common method to solve valuation problems. The model estimates have also been done with the help of GIS procedures that are specific for this kind of analysis.

The model best adjusted to the existing data had an R^2 equal to 0.81 and its estimated parameters are shown in table 3, together with the results of the t test. The model presented errors with Normal Distribution and constant variance. The values of the t Statistics were compared to tabled values, with a 5% confidence level. Examination of these results shows that the non-significant values are only: intercept, location in the block, fence and the interaction factor between area and width. Consequently, from the results of the tests, it can be stated that these variables do not significantly influence the value of a square meter of land. As a result of this method, the mean relative estimation error was 18.45, which is about one-third of the total relative errors of the first method. Once again, the GIS software proved its usefulness for the application.

5 – MODEL 3: ARTIFICIAL NEURAL NETWORKS

Finally, the performance of Artificial Neural Networks is tested for this kind of application. The software employed here is the 4.52 version of a commercial brand named Neural Planner. With this software we use the MLP (Multlayer Perceptron) topology and standard backpropagation learning algorithm (Haykin 1994; Beale and Jackson 1990). The number of hidden layers and the number of neurons in each of these layers are calculated by simulation.

Initially, the data was not subjected to any kind of pre-processing and several networks were built for different combinations of learning rates, *momentum* and number of neurons in the hidden layer. Because the results were not as satisfactory as expected, normalization of the numerical data between 0 and 1 was done. The results continued to be unsatisfactory, however, so another normalization, between 0.1 and 0.9, was done. Different learning rates, *momentum* and number of neurons in the hidden layer were also tested. Table 4 presents the best results.

The values displayed in table 4 show that the network with data normalized between 0.1 and 0.9 presented the best results. It can also be observed that the best values to learning rate and *momentum* are 0.3 and 0.4. Because most properties have neither fences nor paved sidewalks, these variables were removed from the input data and new networks were trained without them, resulting in decreased errors, as shown in table 5.

Table 3: **Regression model adjusted to the available data.**

VARIABLE	ESTIMATED PARAMETERS	t STATISTICS	RESULT OF SIGNIFICANCE TEST
Intercept	4.18	0.66	Non-significant
Distance to the city center (DC)	-30.71	-8.86	Significant
Shape	-4.52	-3.57	Significant
Location	2.26	1.33	Non-significant
Fence	-5.51	-2.25	Significant
Paved Sidewalk	-0.45	-0.08	Non-significant
$(Area)^{-0.5}$	674.13	9.72	Significant
$(Width)^{-0.6}$	66.53	2.96	Significant
DC^2	8.14	6.84	Significant
Area times Width	-1.20	-1.13	Non-significant
Area times DC	0.0013	3.59	Significant
Width times DC	0.08	2.17	Significant

Table 4: **Results of the best ANNs simulations using all the variables.**

MODEL	**NODES IN THE HIDDEN LAYER**	**LEARNING RATE**	***MOMENTUM***	**MEAN**
Raw data	3	0.3	0.3	28.84
Raw data	7	0.5	0.8	28.79
Raw data	4	0.5	0.8	31.99
Raw data	7	0.3	0.8	30.65
Raw data	7	0.2	0.8	30.02
Raw data	7	0.4	0.4	27.48
Raw data	7	0.3	0.3	29.48
Normalized 0 - 1	7	0.4	0.3	29.29
Normalized 0 - 1	7	0.5	0.8	31.08
Normalized 0.1 - 0.9	3	0.3	0.4	33.22
Normalized 0.1 - 0.9	7	0.3	0.4	17.47
Normalized 0.1 - 0.9	4	0.3	0.4	16.91

Some other variables were removed in an attempt to improve the results, but this caused no further reduction of errors. Therefore, the chosen network was the one with two neurons on the hidden layer, using a learning rate equal to 0.3, a *momentum* equal to 0.4 and without the variables fence and paved sidewalks. This network produced a total relative error equal to 14.63.

The results obtained with the ANNs showed a significant reduction of errors in comparison to the other two previous methods. The error found with the ANNs is about one quarter of the error observed with the first model and is roughly twenty-six percent smaller than the error produced by the regression model.

Table 5: **Results of the best ANNs simulations without the variables fence and paved sidewalks.**

MODEL	**NODES IN THE HIDDEN LAYER**	**LEARNING RATE**	***MOMENTUM***	**MEAN**
Normalized 0.1 - 0.9	2	0.3	0.4	14.63
Normalized 0.1 - 0.9	3	0.3	0.4	14.99

6 CONCLUSIONS

This work aimed to study the efficacy of three valuation methods used to estimate urban land values. In the first one, a very simple method, the value of a square meter of land is calculated as a result of an increase or decrease in the value of a standard land parcel. Several aspects, such as size variations, location of the lot, etc may cause variations in value. In Brazilian cities, where this valuation method is quite commonly employed, the variables usually considered are lot length and area. The first valuation method tested in this study tried to reproduce the same approach used in Brazilian cities, but it considered several other characteristics (such as shape and transportation accessibility) in order to make it more reliable. Nonetheless, the weights have been adopted here in the same way that they are in small and medium-sized Brazilian cities. They are mainly based on an individual's decision, usually the town's mayor, following a previous analysis and one or two market surveys and interviews conducted by his staff. The errors (60.48) found with this valuation method were quite high even considering that they are relative errors and that the method considered several variables instead of only the traditional lot dimensions.

The second method studied was the land value estimate based on multiple regression techniques. A significant reduction in errors was found with this method, i.e. from 60.48 in the first case to 18.45 in this second one. Although the results of the regression model for Araçariguama were good, with significantly reduced differences between real and estimated values, there was still room for improvements. We were able to prove this after testing the third method studied here: the estimate based on Artificial Neural Networks. Using this method, the relative errors were reduced to 14.63, representing a reduction of 26.11% when compared to the results of the multiple regression model. Considering these findings, Artificial Neural Networks appear to be a very interesting alternative for real estate valuation, offering much more accurate estimates than traditional methods.

The analyses conducted in this study were supported by a GIS software, which was used to store and make calculations with the data, and to display the results. The GIS built-in tools were able to perform all the calculations for the first and second methods described in this article, without requiring transference of data to any other program. As an additional advantage, it was also able to calculate the shortest route from the city's downtown area to each one of the studied lots. The third method, however, could not make use exclusively of GIS resources because it has no specific built-in ANNs procedure. In the case of this study, a hard interface was built to solve this problem. The results of the tests conducted here indicate the advantages of integrating this kind of routine directly into the GIS program as a first step for a Decision Support System. This would allow users to benefit from the potential of both GIS and ANNs.

ACKNOWLEDGMENTS

The authors would like the express their gratitude to the Brazilian agencies FAPESP - Foundation for the Promotion of Science of the State of São Paulo, CAPES, and Brazilian National Council for Scientific and Technological Development (CNPq), which have supported this research in different manners.

REFERENCES

Abelson, P. (1997) House and Land Prices in Sydney from 1931 to 1989, *Urban Studies* **34**, pp. 1381-1400.

Almond, N.; Jenkins, D.; Gronow, S. (1997) Development of a Prototype Residential Valuation System. *Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management*, Bombai, India, December 16-19, 1997, pp. 204-215.

Aoki, Y.; Osaragi, T.; Ishizaka, K. (1994) An Interpolating Model for Land-price Data with Transportation Costs and Urban Activities. *Environment and Planning* **21**, pp. 53-65.

Azar, K. T.; Ferreira Jr, J.; Abed J.; et al. (1994) Using GIS for Estimating the Spatial Distribution of Land Value in Metropolitan Beirut. *Proceedings of the Conference on Urban and Regional Information Association - URISA.*

Beale, R. and Jackson, T. (1990) *Neural Computing: an Introduction.* Institute of Physics Publishing. Hilger, Bristol.

Bentick, B. L. (1996) The Differential Incidence of the Urban Land Tax Depends on the Travel Intensities of Substitutes for Land. *Urban Studies* **33**, pp. 1729-1732.

Borges, J. M. M. (1975) Custo de Transporte e Valor dos Terrenos Urbanos. MSc Thesis. Faculty of Business and Management, University of São Paulo, São Paulo.

Brondino, N. C. M. and Silva, A. N. R. (1997) Comparando Estratégias de Avaliação de Imóveis para Fins de Tributação com o Auxílio de um SIG. *Proceedings of the Conference GIS Brasil 1997*. Curitiba, Brazil, May 12-16, 1997, CD-ROM.

Draper, N. R. and Smith, H. (1981) *Applied Regression Analysis.* Wiley Series in Probability and Mathematical Statistics, New York. Second Edition.

Falas, T. (1995) Neural Networks in Empirical Accounting Research: An Alternative to Statistical Models. *Neural Network World* **5**, pp. 419-432.

Gazeta de Araçariguama (1995) *Suplemento Especial*, 30 de dezembro de 1995.

Gibb, K (1994) Before and After Deregulation: Market Renting in Glasgow and Edinburgh. *Urban Studies* **31**, pp. 1481-1495.

González, M. A. S. and Formoso, C. T. (1995) Integração dos Tributos Imobiliários Urbanos: Uma Planta de Valores Inferencial Única, Determinada com Dados do ITBI. *Proceedings of the 8th Congresso Brasileiro de Engenharia de Avaliações e Perícias*, Florianópolis, Brazil, October, 1995, pp. 326-335.

Guedes, J. C. (1995) O Emprego de Inteligência Artificial na Avaliação de Bens. *Proceedings of the 8th Congresso Brasileiro de Engenharia de Avaliações e Perícias*, Florianópolis, Brazil, October, 1995, pp. 368-374.

Haykin, S. (1994) *Neural Networks: A Comprehensive Foundation.* Macmillan College Publishing Company, New York. pp. 1-88.

Kwon, O.; Golden, B.; Wasil, E. (1995) Estimating the Length of the Optimal TSP Tour: An Empirical Study Using Regression and Neural Networks. *Computers & Operations Research* **22**, pp. 1047-1056.

Lim, D. (1992) The Nonneutrality of the Land Value Tax: Impacts on Urban Structure. *Journal of Urban Economics* **32**, pp. 186-194.

Möller, L. F. C. (1995) *Planta de Valores Genéricos – Avaliação de Imóveis Para Fins Tributários.* Sagra – DC Luzzatto Editores. Porto Alegre, Brazil. pp. 1-79.

Pasha, H. A. and Butt, M. S. (1996) Demand for Housing Attributes in Developing Countries: A Study of Pakistan. *Urban Studies* **33**, pp. 1141-1154.

Phang, S. Y. and Wong, W. K. (1997) Government Policies and Private Housing prices in Singapore. *Urban Studies* **34**, pp. 1819-1829.

Raia Jr., A. A.; Silva, A. N. R.; Lima, R. S (1996). Utilizando um SIG Para Avaliar Níveis de Acessibilidade em Uma Cidade Média. *Proceedings of the 2nd Congresso Brasileiro de Cadastro Técnico Multifinalitário*, Florianópolis, Brazil, October 13-17, 1996, pp. 193-204.

Subramanian, V.; Hung, M. S.; Hu, M. Y. (1993) An Experimental Evaluation of Neural Networks for Classification. *Computers & Operations Research* **20**, 1993, pp. 769-782.

Tubb, N. R. (1993) A Development Path to Success in Neural Computing. *Expert Systems Applications* **9**, pp. 5-9.

Vertelo, J. B. (1996) Planta Genérica de Valores da Cidade de Governador Valadares. *Proceedings of the 2nd Congresso Brasileiro de Cadastro Técnico Multifinalitário*, Florianópolis, Brazil, October 13-17, 1996, p. 232-243.