A Geostatistical Modeling of Urban Land Values in Milwaukee, Wisconsin

Jun Luo and Yehua Dennis Wei

Department of Geography, University of Wisconsin at Milwaukee
Milwaukee, WI 53201, U.S.A

Abstract

Most studies on urban land values focus on the determinants, using hedonic land price models with the consideration of the physical and socio-economic factors. This article employs geostatistical methods to analyze urban land values. Through a case study of Milwaukee, we set up semivariogram models and 3-D TIN surface models for urban land values, and explore anisotropy characteristics and the relationships of land values among different types of land use. We have found that spatial dependency is a salient feature of urban land values, and spatial clustering of land values varies with type and location of land use. Our models can detect and well explain spatial autocorrelations of urban land values. We demonstrate that geostatistical methods have a great potential when applied in the urban context.

I. INTRODUCTION

Quantitative studies of urban spatial structure have been conducted mainly through analyzing population distribution and density (Bourne, 1989; Batty and Kim, 1992), housing and land prices (Peiser, 1987), and firm location (Shukla and Waddell, 1991). Such studies often require a substantial amount of data from population census, records of property transactions, and firm-level data. Housing price and land value models have often been used to study urban spatial structure and urban economies, and are central to the debate on race and equity (Kim, 2000).

The uneven distribution of urban land values has attracted considerable scholarly attentions (Abelson, 1997; Bertaud, 1992; Brigham, 1965; Huh and Kwak, 1997; McDonald and McMillen, 1998; McMillen, 1990, 1996; McMillen and McDonald, 1991). Many of the studies deal with the determinants of land values using hedonic land price models, with the consideration of physical and socio-economic factors, such as race, housing attributes, and neighborhood. Regression models have been developed to explain land values and housing prices with a number of independent variables (Erickson, 1986; Peiser, 1987). To simplify the models, many researchers focus on a singular determinant of land values, a singular type of urban land use, and/or the distance decay of urban land values.

Previous studies of urban land values, however, have several limitations. First, despite the change from a monocentric to a polycentric structure of cities, traditional negative exponential density function is still widely used to model land values and urban spatial structure. Those models are over-simplified and therefore are less effective in examining spatial differences in land value distribution. Second, many studies on urban spatial structure primarily use aggregate zonal data. There are problems associated with this form of data, such as "Modifiable Areal Unit Problem"(Openshaw, 1984); findings vary with the change in the level of aggregation and the configuration of the zoning system. Lastly, spatial statistics have great potential to improve the understanding of urban land use, housing markets, and urban changes, but its application in modeling land values remains limited. Previous researches have been hampered by poor data and the limited usage of spatial statistics. Housing markets exhibit spatial dependency, and almost all hedonic models violate regression assumptions (Mulligan et al., 2002). More studies of the spatial distribution of urban land values and the relationship among different types of urban land uses are needed.

Consequences of urban restructuring and extents of spatial segregation remain hotly debated issues for Milwaukee. While some have argued for the decline of residential segregation in the city, many others have maintained that segregation persists and inner-city neighborhoods are troubled by unemployment and poverty (e.g., Boardman and Field, 2002). Regarding land values, many studies have focused on cities like Chicago (e.g., McMillen, 1996), and few studies have examined Milwaukee. While Kim (2000) examined the relationship between race and home price appreciation by neighborhoods, no study has examined the spatial patterns of land values and the effects of residential segregation in the city.

In this paper, we use geostatistical approaches to study urban land values, through a case study of the City of Milwaukee. The paper attempts to examine the characteristics of spatial distribution of urban land values, to develop models for urban land value distribution with kriging methods, and to analyze their implications for urban spatial structure. We stress the feature of spatial dependency and the clustering of urban land

1082-4006/04/1001-49$5.00

values, and argue that race remains a critical determining factor in accounting for the uneven distribution of land values in the Midwest cities. The paper is organized as follows: after the next section on data and methodology, we set up and interpret land value distribution models. Then we analyze the implications of land value models for urban spatial structure, and the last section is the conclusion.

II. DATA AND METHODOLOGY

Milwaukee is a classic rustbelt city, which has experienced the decline of manufacturing jobs and is one of the most segregated cities in the United States. In 2000, the city had a population of 596,974, considerably less than the 717,372 of 1970 (Huang and Wei, 2002). African-Americans are heavily concentrated in the inner-city north, followed by the northwest areas, while the suburban areas in northeast, west, and south have a much lower share of African American population (Figure 1).

Data for this research is from the Master Property File (MPROP) datasets created by the City Government of Milwaukee. MPROP is a computerized inventory of all properties in the City of Milwaukee. It contains more than 90 indicators describing each of the approximately 160,000 properties in the city. The file was firstly created in 1975 to provide current and accurate property information and can be accessed in a variety of ways. Our research mainly uses the 2001 dataset since it was the most recent data available.

Grid sampling method of geostatistics is used in the research (Wackernagel, 1998). We create a fishnet, which covers the whole area of the city. The cell size is 150 feet because a bigger cell size may not reflect the actual change of spatial characteristics, while a smaller size may make one sample lie across more than one cell. We select at most one sample property in each cell. We have 526 samples, including 329 residential samples, 100 commercial samples, and 97 manufacturing samples; each of them is associated with their assessed land values in 2001. Figure 2 presents the spatial distribution of the research samples.

To analyze spatial distribution and clustering, especially spa-

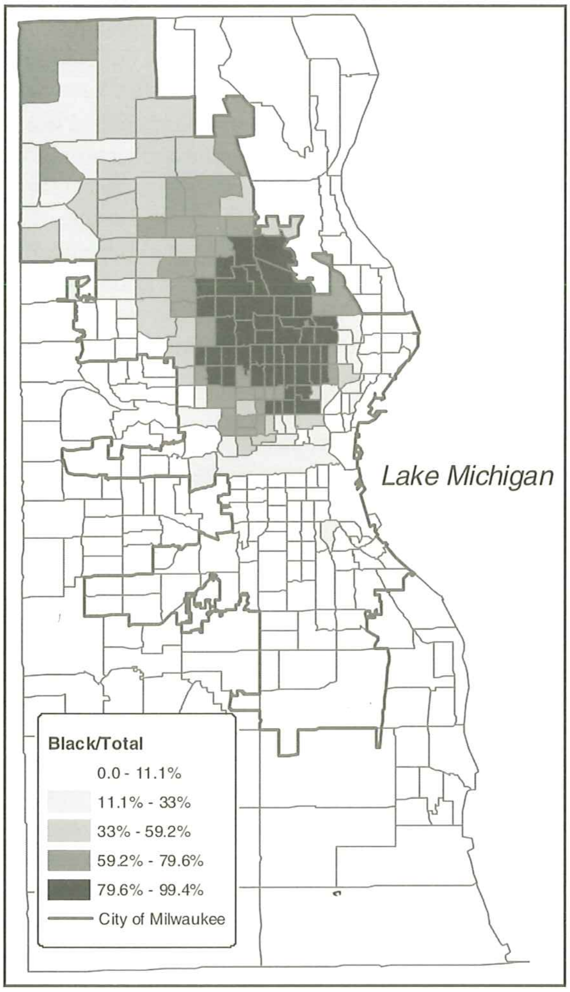


Figure 1. Distribution of African-American population in the city of Milwaukee, 2000.

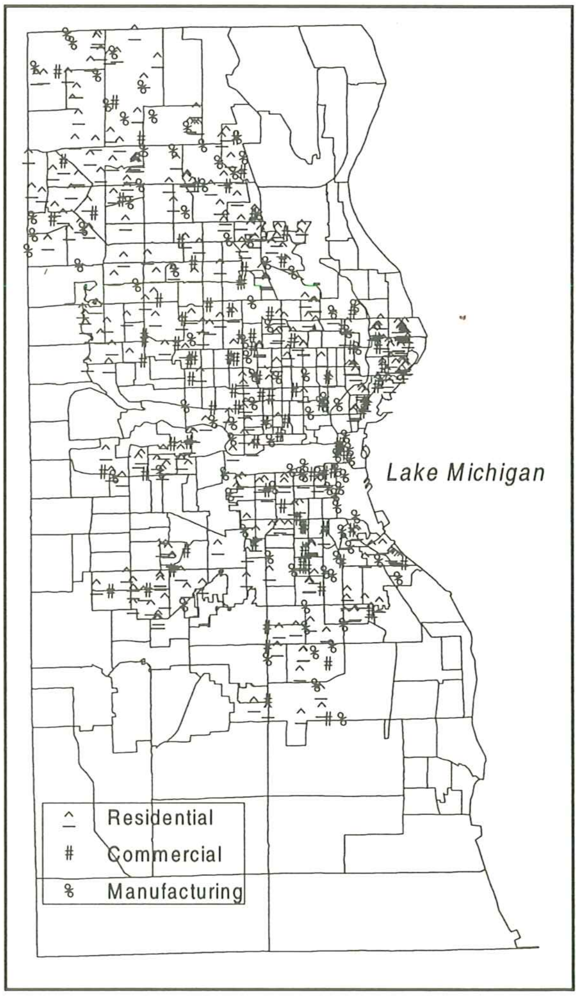


Figure 2. Spatial distribution of research samples.

tial autocorrelation, we use the global Moran's I statistic, which is calculated from the following formula:

$$I = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} z_i z_j}{(N-1)\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \qquad (1)$$

where $z = x - \bar{x} / s$ is the z-score of the variable of interest x; $w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{N} c_{ij}}$, c_{ij}=1 if i is within a critical distance to j, and $c_{ij} = 0$ otherwise, w_{ij} then forms a row-standardized weights matrix.

The research adopts the geostatistical method, mainly semivariogram models and kriging, which are based on statistical models that include autocorrelation (statistical relationships among the measured points). Assuming we have a population of samples in space, (S_1, S_2, ...S_n), and their attributes, (Z_1,Z_2,......Z_n), empirical semivariogram for the (i,j)th pair can be calculated by the following:

$$0.5 * (Z(S_i) - Z(S_j))^2 \qquad (2)$$

in which $Z(S_i)$ and $Z(S_j)$) are the values in the locations S_i and S_j respectively. After calculating all pairs' semivariogram, we can bin the semivariograms based on common distance and direction:

$$r(d) = \frac{1}{2N(d)}\sum_{i=1}^{N}[Z(s_i) - Z(s_i - d)]^2 \qquad (3)$$

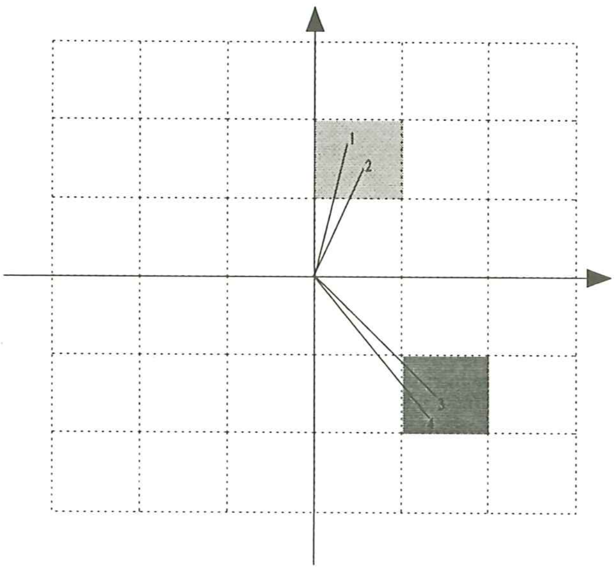


Figure 3. Grid tolerance regions for binning empirical semivariograms.

in which a bin $r(d)$ is the half of the averaged sum of the squared difference from the values for all pairs of locations with common distance and direction. The most commonly used binning method is tolerance regions that are rectangles and distributed uniformly on a grid (Figure 3). The cell size of the grid is also called lag size. After plotting the empirical semivariogram value for each bin for each direction, we get the empirical semivariogram cloud.

After estimating the empirical semivariogram, we fit a theoretical model to the empirical semivariogram. The most commonly used theoretical models include spherical, exponential, tetraspherical, pentaspherical, and gaussian. These theoretical semivariogram models are based on the intrinsic stationarity assumption that the variance is the same between any two points that are at the same distance and direction apart no matter which two points you choose. After testing these commonly used models with the cross-validation method (see section IV), we adopt the nugget spherical semivariogram model, which is presented in the following form:

$$r(h;\theta) = Nugget + \begin{cases} \theta_s \left[\frac{3}{2}\frac{\|h\|}{\theta_r} - \frac{1}{2}\left(\frac{\|h\|}{\theta_r}\right)^2 \right] & for \quad 0 \le \|h\| \le \theta_r \\ \theta_s & for \quad \theta_r \le \|h\| \end{cases} \qquad (4)$$

where h is the distance, $\theta_s \ge 0$ is the partial sill parameter and $\theta_r \ge 0$ is the range parameter. The sill, nugget, and range are important parameters when fitting a theoretical model to empirical semivariogram (Figure 4). The partial sill is the sill minus nugget. We expect that the semivariogram functions change not only with distance but also with direction. This is called anisotropy, or directional semivariograms. In general, anisotropy means that the ranges vary with the directions. We need to consider anisotropy effect when deciding the search neighborhood size in kriging prediction.

Kriging is the interpolation method of geostatistics. Generally, kriging interpolation takes the following form:

$$\hat{Z}(0) = \sum \lambda_i Z(x_i), \qquad \sum \lambda_i = 1 \qquad (5)$$

where $Z(0)$ is the attribute value of unmeasured point in space. λ_i can be solved through the following kriging equations:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ m \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix} \qquad (6)$$

where m is the lagrangian multiplier. At a certain distance, the sample points have no correlation with the prediction location, and it is possible that they may even be located in an area very different than the unknown location. Determining the shape and size of search neighborhood is a complex process,

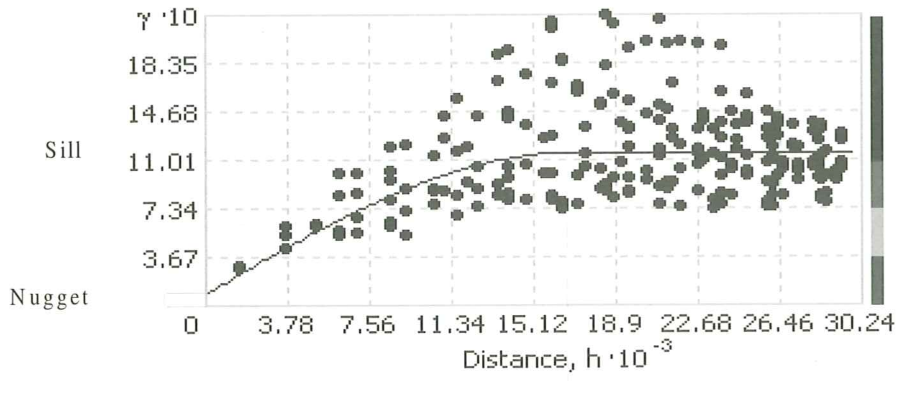
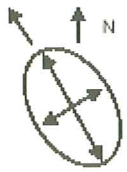


Figure 4. Empirical semivariogram clouds and isotropic fitting model. Note: Distance h in foot.

which requires a good understanding of the sample data. An example of search neighborhood is given as the below:



where the long axis of the ellipse is the major range and the short axis is the minor range with each one's direction in two dimensional space.

We use the Gstat package, which provides powerful tools for geostatistical modeling. The package improves the speed and efficiency of the computing process, but the decision on parameters requires a careful and comprehensive analysis.

III. MODELING URBAN LAND VALUES

Table 1 presents results of spatial autocorrelation analysis (as normalized values). We can see that land values have a high degree of positive spatial association, indicating that high values tend to cluster with high values, and vice versa. Residential land values have the highest spatial association, followed by commercial and manufacturing land values. These findings provide further evidence for the high level of residential segregation in Milwaukee.

Table 1. Global spatial autocorrelation analysis

Land Value	Standardized Moran's *I*
Overall	25.642**
Commercial	24.357**
Residential	32.189**
Manufacturing	12.794**

** indicates significant at the 5 per cent level

We model the spatial distribution of urban land values in the city as a whole using the isotropic model and the anisotropic model. Figure 4, where the y-axis is the empirical semivariogram value and the x-axis is the distance associated with the bin, presents the isotropic fitting model. The model is calculated as follows:
(see equation (7) on next page)

Figure 4 shows that land value differences increase with the increase of distance between measured observations. Beyond the distance of 3.19 miles (16823 feet) as the range parameter, land value differences become constant and measured points are spatially uncorrelated. This suggests that land values are spatially associated only at locations within certain distance.

Directional autocorrelation (i.e. anisotropy) can be examined by considering various search directions. Figure 5 shows the fitting models in various search directions and suggests that with different directions land value differences may vary even though they are at the same distance.

We then extract the highest and lowest semivariogram curves from Figure 5. Figure 6 and Figure 7 are semivariograms in directions NNW and WSW with an angle of 345.2 degree and

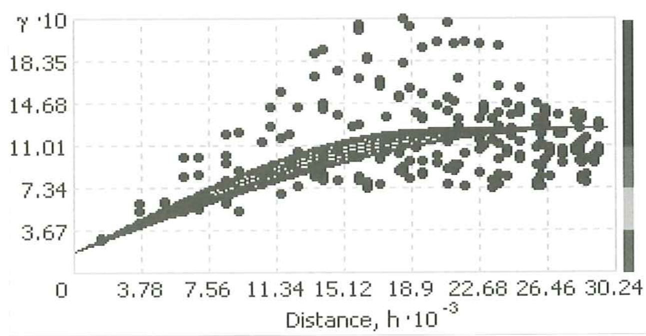


Figure 5. Fitting models in various directions. Note: Distance h in foot.

$$r(h;\theta) = 0.090028 + \begin{cases} 1.0646\left[\frac{3}{2}\frac{\|h\|}{16823} - \frac{1}{2}\left(\frac{\|h\|}{16823}\right)^2\right] & for \quad 0 \le \|h\| \le 16823 \\ 1.0646 & for \quad 16823 \le \|h\| \end{cases} \tag{7}$$

$$r(h;\theta) = 0.17657 + \begin{cases} 1.0687\left[\frac{3}{2}\frac{\|h\|}{28458} - \frac{1}{2}\left(\frac{\|h\|}{28458}\right)^2\right] & for \quad 0 \le \|h\| \le 28458 \\ 1.0687 & for \quad 28458 \le \|h\| \end{cases} \tag{8}$$

$$r(h;\theta) = 0.17657 + \begin{cases} 1.0687\left[\frac{3}{2}\frac{\|h\|}{20368} - \frac{1}{2}\left(\frac{\|h\|}{20368}\right)^2\right] & for \quad 0 \le \|h\| \le 20368 \\ 1.0687 & for \quad 20368 \le \|h\| \end{cases} \tag{9}$$

65.2 degree respectively.

As shown in Figure 6, in the NNW direction land value differences increase slowly with distance, and become constant beyond the distance of 5.39 miles. In the WSW direction, however, land value differences increase faster and turn to constant at the distance of 3.91 miles (Figure 7). The influence of anisotropy is apparent, since in the direction of NNW, land values are spatially correlated in a bigger range than that in the direction of WSW. We therefore use them to determine the search neighborhood's shape and size.

In the NNW direction with the range of 5.39 miles (28458 feet) and the angle of 345.2 degree, an anisotropy fitting model is constructed as follows:
(see equation (8) above)

In the direction WSW with the range of 3.91 miles (20368 feet) and the angle of 65.2 degree, the fitting model becomes:
(see equation (9) above)

Based on the above two directional fitting models, we can determine the searching neighborhood's shape and size as Figure 8. Due to the distribution of samples, we have many more samples in the NNW direction than in the WSW direction, which makes the search neighborhood's major and minor axes somewhat consistent with the city shape. The length of semi-major axis is 5.39 miles and the length of semi-minor axis is 3.91 miles. Since our sample data are collected on a grid, we divide the ellipse into four sectors; each sector has two to five measured points, which are used for further analysis.

Based on the above kriging equation's 5 and 6, and the searching neighborhood, we interpolate all unmeasured points in the city. A grid is generated to present the land value surface of Milwaukee City (Figure 9). The land value surface of Milwaukee City manifests that the highest land values cluster in the lakefront and downtown areas, while the lowest land values cluster in inner-city northern area.

Furthermore, a TIN is created based on the land value surface grid, which provides a 3-D view of the spatial distribution of land values (Figure 10). Similarly, the TIN shows that land values peak in the lakefront and downtown areas, and the

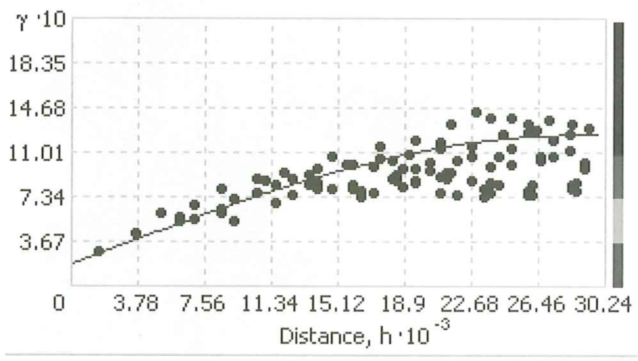


Figure 6. Semivariogram cloud in the direction of NNW. Note: Distance h in foot.

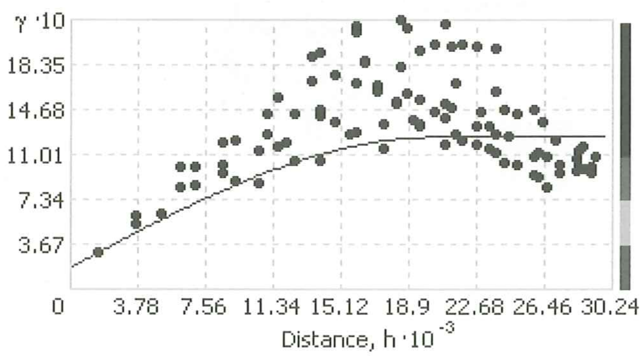


Figure 7. Semivariogram cloud in the direction of WSW. Note: Distance h in foot.

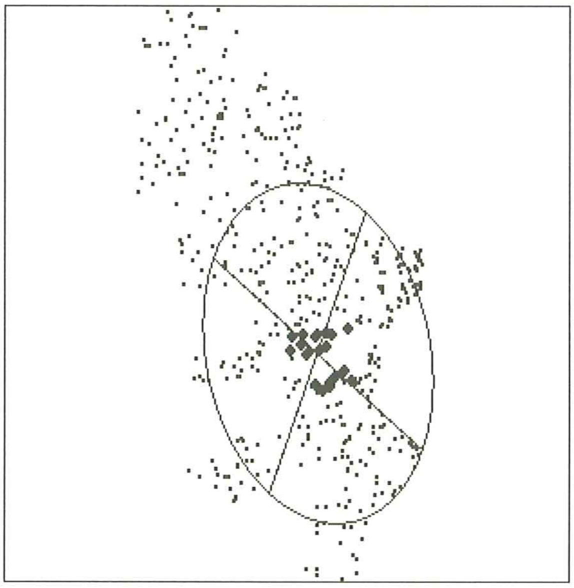


Figure 8. Searching neighborhood for the overall model.

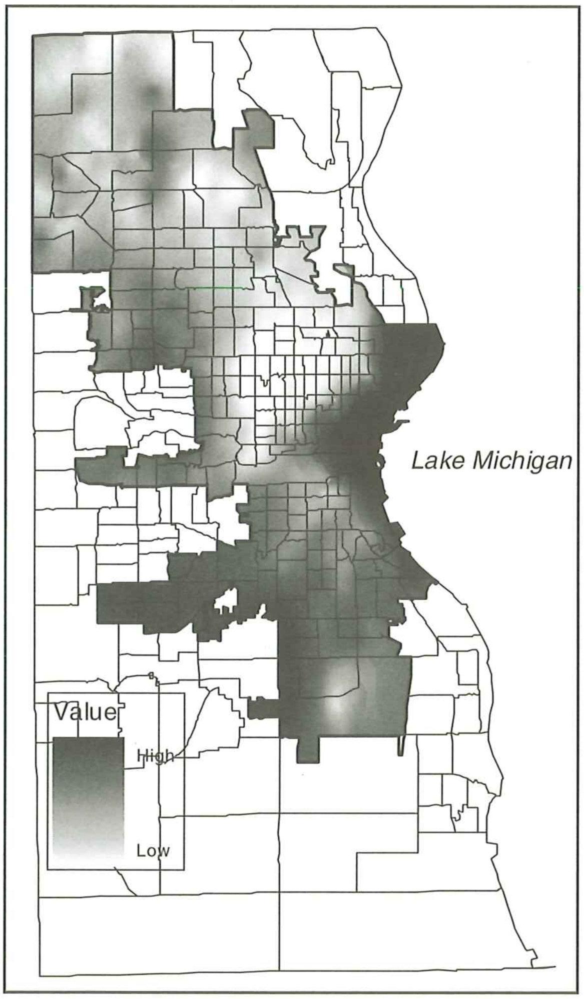


Figure 9. Land value surface in Milwaukee.

land value basin is in the inner city north. Areas in the northwest also have low land values, while the land value surface becomes higher in southwest area.

IV. URBAN LAND USE AND LAND VALUES

Since the above isotropic and anisotropic models show the land values of the city as a whole regardless of the land use type, we further model land value distribution for different land use types. Table 2 presents the parameters of isotropic models, and Table 3 presents the parameters of anisotropic models. We have also created three TINs to show the land value distribution of residential, commercial and manufacturing land uses.

Results of isotropic modeling presented in Table 2 indicate that commercial land values have the highest aggregation level (indicated by partial sill), followed by manufacturing land values and residential land values. This implies that in general, commercial land use values are the most centralized, with higher values in the downtown area and decreasing quickly with the increase of distance (Figures 11, 12, and 13). The areas with the lowest land values are not consistent with the areas with the largest distances to the downtown, but in the inner-city north and old near suburban areas in the southeast.

Residential land use is significantly decentralized, and does not simply follow the distance decay model. The areas with high land values are in the downtown areas and the lakefront areas in the northeast, followed by the areas in the southwest

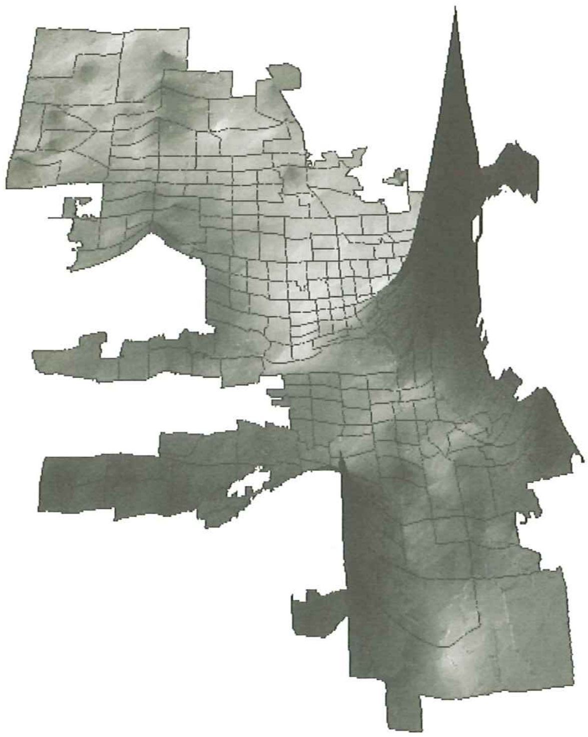


Figure 10. Land value surface TIN in Milwaukee.

Table 2. Parameters of isotropic models

Parameter	Nugget	Partial Sill	Range(foot)
Overall	0.090028	1.0646	16823
Residential	0.89869	8.2221	17572
Commercial	0	18.101	13485
Manufacturing	0	17.94	6009.3

Table 3. Parameters of anistropy models

Parameter		Nugget	Partial Sill	Range(foot)
Overall	NNW(345.2•)	0.17657	1.0687	28458
	WSW(65.2•)	0.17657	1.0687	20368
Residential	NNW(348.4•)	1.6017	8.2826	28377
	WSW(78.4•)	1.6017	8.2826	22864
Commercial	NNW(333.9•)	2.6965	16.927	23608
	WSW(63.9•)	2.6965	16.927	17049
Manafacturing	NNW(297.8•)	0	19.032	13734
	WSW(27.8•)	0	19.032	6048.9

(Figure 11). The areas with low land values are in the inner-city north, and spots in the northwest where blacks have moved in recently. Manufacturing land values are relatively centralized, and the areas with the lowest values are in the inner-city north (Figure 13).

The cross validation method, which removes one sample and uses the rest of the sample data to predict the removed sample, is adopted to determine the prediction accuracy. Table 4 presents the results of cross validation. The standardized means of the four models are all near zero and each model has root-mean-square near the average standard error. This indicates that the predictions of our models are reasonably accurate, and that our geostatistical models of urban land values provide an effective way to investigate urban land values and spatial structure.

Although our land value models are based on the geostatistical analysis rather than the hedonic models, the determining factors are embodied in the land value models. Residential land use has the greatest influence range (indicated by the range parameter), followed by commercial and manufacturing land use. This characteristic shows that residential land use has the strongest effect of agglomeration: high land values tend to cluster and so do the low land values. The limited influence range of manufacturing land value means that the urban manufacturing function is weak and the proportion of manufacturing land use is low in the city, although Milwaukee is traditionally a manufacturing city.

The anisotropic models show that urban land value distribution varies in different directions. In the direction of NNW, all three land use types have the biggest influence range. Residential land use and commercial land use have the similar directional angle, which indicates that decentralization has the similar direction and is relatively synchronous. This further confirms the significance of residential segregation in Milwaukee. The distribution angle of manufacturing land value search neighborhood is different from that of residential and

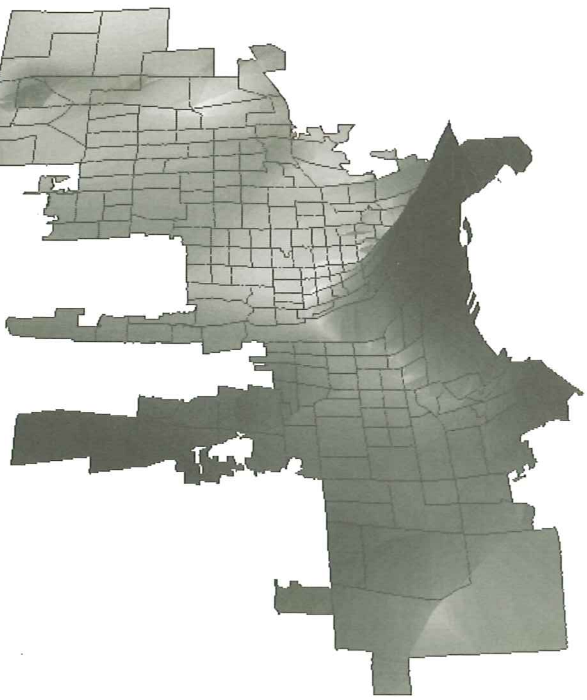


Figure 11. Residential land value surface TIN.

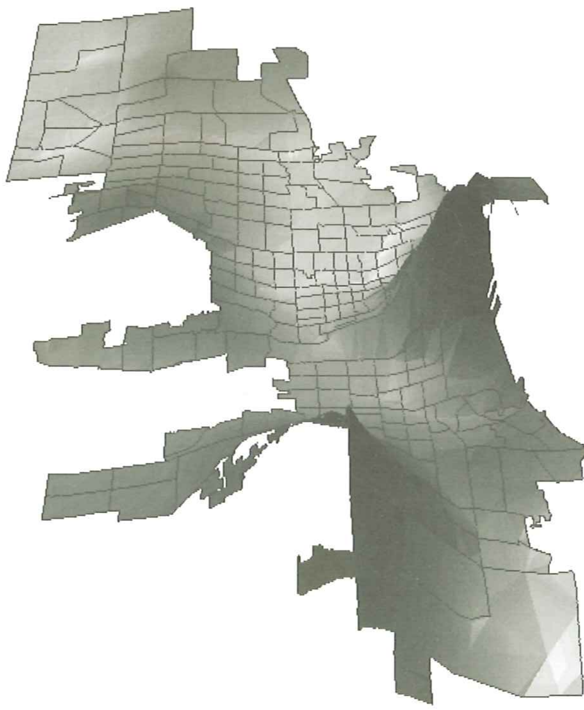


Figure 12. Commercial land value surface TIN.

Table 4. Cross validation results

Predictions Errors	Standardized Mean	Root-Mean-Square	Average Standard Error
Overall	0.01112	1.949	2.112
Residential	0.0008725	1.054	1.656
Commercial	0.001226	1.738	2.817
Manufacturing	-0.05311	4.269	4.744

commercial land values, which is more oriented towards the west, and is pointed to the employment sub-center identified by McMillen (2001), where the Harley-Davidson motor plant is located (see Figure 14).

Land value distribution maps and models clearly reflect residential segregation in Milwaukee. Figure 1 shows that the inner city north area has the largest percentage of African American population, followed by northwest areas, while the suburban areas in the northeast, west, and south have much smaller shares of African American population. Our land value models coincide with and reflect such spatial patterns of population distribution. Neighborhoods with a large percentage of the minority population, such as the inner city north and northwest areas, have low land values.

As presented in the four land-value surface TINs, the downtown and the lakefront areas have the highest land values while the inner city north has the lowest land values. It can also be concluded that suburbanization and agglomeration levels are relatively higher in the southwest fringe area than other city fringe areas. Land values reach another peak in the southwest fringe area, although they are still much lower than the downtown and lake front areas (Figure 9).

V. CONCLUSION

Geostatistical methods have been well developed and are often used in natural sciences such as geology. This study uses geostatistical methods to model the spatial distribution of urban land values. The spatial distribution of urban land values is often influenced by socio-economic factors. Semivariogram models have been proven to be effective when modeling the spatial autocorrelations in urban land values, but weak when reflecting the social-economic factors influencing the spatial distribution of urban land values. Moreover, kriging interpolation only uses the location information to predict the places with unknown land value. Cross-semivariogram, however, can use a secondary variable such as population distribution to model the spatial distribution of urban land values as primary variable. More importantly, cross-

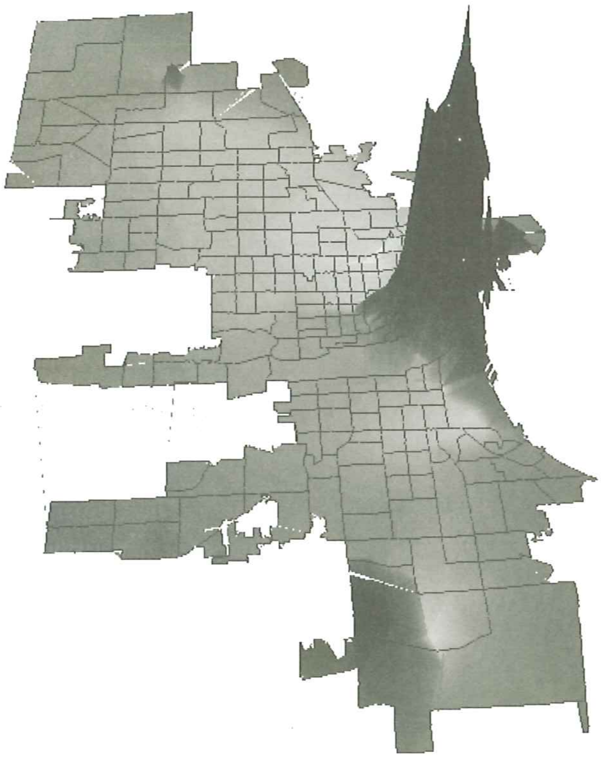


Figure 13. Manufacturing land value surface TIN.

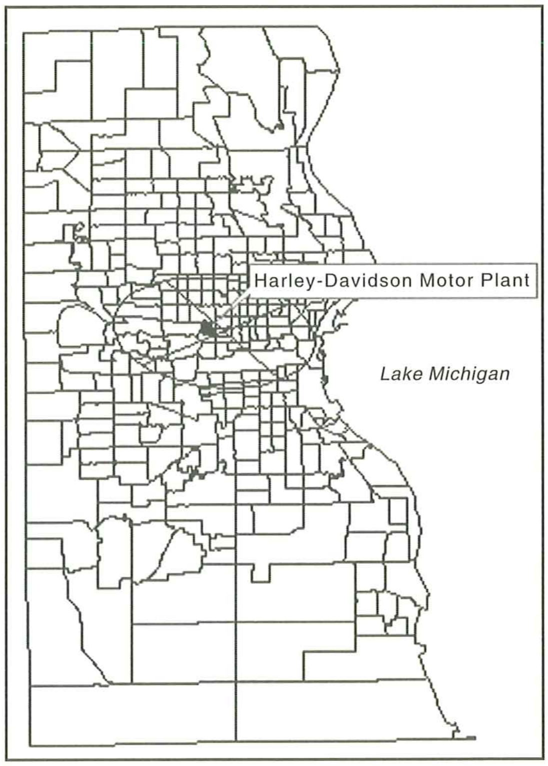


Figure 14. Manufacturing land value search neighborhood.

semivariogram can reveal the spatial relationship between the primary and secondary variables. Cokriging is based on the cross-semivariogram and implements the interpolations using both the location and socio-economic factors, which makes the prediction more accurate. Future researches should try to incorporate the socio-economic variables into cross-semivariogram models, and make the geostatistical modeling more powerful to explain spatial patterns of urban land values.

In this study, we have used geostatistical methods in a GIS environment to study urban land values, and demonstrated their utilities when applied in the urban context. Unlike the commonly used regression models based on hedonic land price function, our geostatistical models focus on the spatial dimensions of urban land values and can be used effectively in investigating spatial patterns of land values and urban spatial structure.

Through a case study of Milwaukee, we have developed isotropic and anisotropic semivariogram models for urban land values of different land-use types, and performed point interpolations using the kriging methods. We also have created 3-D TIN to visualize spatial patterns of land values. Our models have shed some light on urban spatial structure through analyzing the parameters of geostatistical models of land values. In particular, we have revealed the spatial differentials of the extent of agglomeration among residential, commercial and manufacturing land uses. Our models have profound implications for urban social structure and can be used as an effective way to investigate urban social space when combined with socio-economic data such as population.

REFERENCES

[1] Abelson, P., 1997,House and Land prices in Sydney from 1931-1989, *Urban Studies*, 34:1381-1400.

[2] Batty, M and Kim, K.S., 1992, From follows function: reformulating urban population density functions, *Urban Studies*, 29:1043-1070.

[3] Bertaud, A, 1992, The impact of land-use regulations on land supply, consumption, and price, *Regional Development Dialogue,* 13: 35-40.

[4] Boardman, J.D. and S.H. Field, 2002, Spatial Mismatch and Race Differentials in Male Joblessness: Cleveland and Milwaukee, 1990. *The Sociological Quarterly,* 43(2): 237-255.

[5] Brigham, E.F., 1965, Determinants of residential land values, *Land Economics*, 41: 325-334.

[6] Bourne, L.S., 1989, Are new urban dorms emerging? Empirical tests for Canadian urban areas, *Canadian Geographer,* 33: 312-328.

[7] Erickson, R.A., 1986, Multi-nucleation in metropolitan economies, *Annals of the Association of American Geographers*, 76:331-346.

[8] Huang, R.H. and Y.H.D. Wei., 2002, Analyzing Neighborhood Accessibility via Transit in a GIS Environment, *Geographic Information Science*, 8(1):39-47.

[9] Huh, S. and Kwak, S.J.K., 1997, The choice of functional form and variables in the hedonic price model in Seoul, *Urban Studies,* 34:989-998.

[10] Kim, S., 2000, Race and Home Price Appreciation in Urban Neighborhoods:Evidence from Milwaukee, Wisconsin, *Review of Black Political Economy*, 28(2):9-28.

[11] McDonald, J.F. and McMillen, D.P., 1998, Land values, land use, and the first Chicago zoning ordinance, *Journal of Real Estate Finance and Economics*, 16:135-150.

[12] McMillen, D.P., 1990, Consistent estimation of the urban land value function, *Journal of Urban Economics*, 27:285-293.

[13] McMillen, D.P., 1996, One hundred and fifty years of land values in Chicago: a nonparametric approach, *Journal of Urban Economics*, 40:100-124.

[14] McMillen, D.P. and McDonald, J.F., 1991, A simultaneous equations model of zoning and land values, *Regional Science and Urban Economics*, 21:55-71.

[15] McMillen, D.P., 2001, Polycentric urban structure: The case of Milwaukee, *Federal Reserve Bank of Chicago—Economic Perspectives,* 2Q:15-27.

[16] Mulligan, G.F., Franklin, R. and Esparza, A.X., 2002, Housing Prices in Tucson, Arizona, *Urban Geography*, 23(5):446-470.

[17] Peiser, R.B., 1987, The determinants of non-residential urban land values, *Journal of Urban Economics*, 22:340-360.

[18] Openshaw, S., 1984, The *Modifiable Areal unit Problem: Concepts and Techniques in Modern Geography*, Norwich: Geo Books.

[19] Shukla, V. and Waddell, P., 1991, Firm location and land use in discrete urban space: a study of the spatial structure of Dallas-Fort Worth, *Regional Science and Urban Economics*, 21:225-253.

[20] Wackernagel, Hans., 1998, *Multivariate Geostatistic*, New York: Springer.