Modeling Urban Land Values in a GIS Environment

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

May 2004
I. INTRODUCTION

In this paper, I use Geostatistic 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, develop models for urban land value distribution with Kriging methods, and analyze their implications for urban spatial structure. I stress the feature of spatial dependency and the clustering of urban land 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, I set up and interpret land value distribution models. Then I analyze the implications of land value models for urban spatial structure, and the last section is the conclusion.

II. DATA AND METHODOLOGY

Data for this research is from the Master Property File (MPROP) datasets created by the City Government of Milwaukee. Grid sampling method of Geostatistics is used in the research (Wackernagel, 1998). I 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 mercantile samples, and 97 manufacturing samples; each of them is associated with their assessed land values in 2001.

The research adopts the Geostatistical methods, 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, \((S_1, S_2, \ldots, S_n)\), and their attributes, \((Z_1, Z_2, \ldots, Z_n)\), empirical semivariogram for the \((i,j)\)th pair can be calculated by the following:

\[
0.5 \times (Z(S_i) - Z(S_j))^2
\] (1)

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:

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

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.

After estimating the empirical semivariogram, I 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, I adopt the Nugget Spherical semivariogram model, which is presented in the following form:

\[
r(h; \theta) = \text{Nugget} + \left\{ \theta_s \left[ \frac{3}{2} \frac{\|h\|}{\theta_r} - \frac{1}{2} \left( \frac{\|h\|}{\theta_r} \right)^2 \right] \right\} \quad \text{for} \quad 0 \leq \|h\| \leq \theta_r
\]

\[
\text{Nugget} \theta_r \quad \text{for} \quad \theta_r \leq \|h\|
\] (3)
where $h$ is the distance, $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter. The nugget effect can be attributed to the measurement errors or spatial sources of variation at distances smaller than the sampling interval. The sill, nugget and range are important parameters when fitting a theoretical model to empirical semivariogram (Figure 1). 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. I need to consider anisotropy effect when deciding the search neighborhood size in Kriging prediction.

Kriging is the interpolation method of Geostatistic, and depends on models of spatial autocorrelation, which can be formulated as fitted model of semivariograms or covariance. Generally, Kriging interpolation takes the following form:

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

(4)

where $\hat{Z}(0)$ is the attribute value of unmeasured point in space. $\lambda_i$ can be solved through the following Kriging equations:

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

(5)

where $m$ is the Lagrangian multiplier. At 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.
I use the Gstat package, which provides powerful tools for Geostatistic 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

I model the spatial distribution of urban land values in the city as a whole, using the Isotropic model and the Anisotropic model. Figure 1, 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:

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

Figure 1 shows that land value differentials increase with the increase of distance between measured observations. Beyond the distance of 3.19 miles (16823 feet) as the range parameter, the land value difference becomes constant and measured points are not spatially correlated. This suggests that land values are spatially associated only at locations within some certain distance.

Directional autocorrelation (i.e. Anistropy) can be examined by considering various search directions. Figure 2 and Figure 3 are semivariograms in directions NNW and WSW with an angle of 276.2 degree and 6.2 degree respectively.

As shown in Figure 2, in the NNW direction the land value difference among points increases slowly with distance, and becomes constant beyond the distance of 5.39 miles. In the WSW direction, however, the land value difference increases faster and turns to constant at the distance of 3.91 miles (Figure 3). The influence of Anistropy 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 decide the search neighborhood’s shape and size.

In the NNW direction with range 5.39 miles (28458 feet) and angle of 345.2 degree, the Anistropy fitting model is:

\[
\begin{align*}
  r(h; \theta) &= 0.17657 + \\
  &\left\{ 1.0687 \left[ \frac{3}{2} \frac{\|h\|}{28458} - \frac{1}{2} \left( \frac{\|h\|}{28458} \right)^2 \right] \right. \\
  &\quad \left. for \quad 0 \leq \|h\| \leq 28458 \right. \\
  &\quad \left. 1.0687 \right. \quad for \quad 28458 \leq \|h\| \quad (7)
\end{align*}
\]

In the direction WSW with range 3.91 miles (20368 feet) and angle of 65.2 degree, the fitting model is:

\[
\begin{align*}
  r(h; \theta) &= 0.17657 + \\
  &\left\{ 1.0687 \left[ \frac{3}{2} \frac{\|h\|}{20368} - \frac{1}{2} \left( \frac{\|h\|}{20368} \right)^2 \right] \right. \\
  &\quad \left. for \quad 0 \leq \|h\| \leq 20368 \right. \\
  &\quad \left. 1.0687 \right. \quad for \quad 20368 \leq \|h\| \quad (8)
\end{align*}
\]

Based on the above two directional fitting models, we can decide the searching neighborhood’s shape and size as Figure 4. 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 equations and the searching neighborhood, I interpolate all unmeasured points in the city. A grid is generated to present the land value surface of Milwaukee City (Figure 5). 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 north 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 6). Similarly, the TIN shows that land values peak in the lakefront and downtown areas, and the 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.

Since the above models show 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 1 presents the parameters of Isotropic models, and Table 2 presents the parameters of Anistropy models. I have also created three TINs to show the land value distribution of residential, commercial and manufacturing land uses (Figure 7, 8 and 9).

**IV URBAN LAND USE AND LAND VALUES**

The isotropic models presented in Table 1 show that the aggregation level (indicated by partial sill) of commercial land value > manufacturing land value > residential land value. This indicates that in general, compared with other land use types, the commercial land use values are the most centralized, with higher values in the downtown area and decreasing quickly with the increase of distance (Figures 7, 8, and 9). The areas with the lowest land values, however, are not the areas with the largest distances to the downtown, but the inner city north and old near suburban areas in the southeast.

Residential land use is significantly decentralized, and doesn't simply follow the distance decay model. The areas with high land values are the downtown areas and the lakefront areas in the northeast, followed by the areas in the southwest (Figure 7). The areas with low land values are 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 the inner city north (Figure 9).

Cross validation method, which means that one sample is removed and the rest of the sample data is used to predict the removed sample, is adopted to measure the prediction accuracy. Table 3 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 Geostatistic models of urban land values provide an effective way to investigate land values and urban spatial structure.

Although our land value models are based on the Geostatistic analysis rather than the hedonic models, the determining factors are embodied in the land value models. The influence range of urban land values (indicated by the range parameter) can be summarized as the following: the residential land use > commercial land use > 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 manufacturing land use proportion is low in the city, although Milwaukee is traditionally a manufacturing city.

Land value distribution maps and models clearly reflect residential segregation in Milwaukee. Figure 10 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 lake front 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, since land values reach another peak in the southwest fringe area, although they are still much lower than the downtown and lake front areas (Figure 5).

REFERENCES

### Table 1: Parameters of Isotropic Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Nugget</th>
<th>Partial Sill</th>
<th>Range (foot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.090028</td>
<td>1.0646</td>
<td>16823</td>
</tr>
<tr>
<td>Residential</td>
<td>0.89869</td>
<td>8.2221</td>
<td>17572</td>
</tr>
<tr>
<td>Commercial</td>
<td>0</td>
<td>18.101</td>
<td>13485</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0</td>
<td>17.94</td>
<td>6009.3</td>
</tr>
</tbody>
</table>

### Table 2: Parameters of Anisotropy Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Nugget</th>
<th>Partial Sill</th>
<th>Range (foot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNW(345.2°)</td>
<td>0.17657</td>
<td>1.0687</td>
<td>28458</td>
</tr>
<tr>
<td>WSW(65.2°)</td>
<td>0.17657</td>
<td>1.0687</td>
<td>20368</td>
</tr>
<tr>
<td>Residential</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNW(348.4°)</td>
<td>1.6017</td>
<td>8.2826</td>
<td>28377</td>
</tr>
<tr>
<td>WSW(78.4°)</td>
<td>1.6017</td>
<td>8.2826</td>
<td>22864</td>
</tr>
<tr>
<td>Commercial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNW(333.9°)</td>
<td>2.6965</td>
<td>16.927</td>
<td>23608</td>
</tr>
<tr>
<td>WSW(63.9°)</td>
<td>2.6965</td>
<td>16.927</td>
<td>17049</td>
</tr>
<tr>
<td>Manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNW(297.8°)</td>
<td>0</td>
<td>19.032</td>
<td>13734</td>
</tr>
<tr>
<td>WSW(27.8°)</td>
<td>0</td>
<td>19.032</td>
<td>6048.9</td>
</tr>
</tbody>
</table>

### Table 3: Cross Validation Results

<table>
<thead>
<tr>
<th>Predictions Errors</th>
<th>Standardized Mean</th>
<th>Root-Mean-Square</th>
<th>Average Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.01112</td>
<td>1.949</td>
<td>2.112</td>
</tr>
<tr>
<td>Residential</td>
<td>0.0008725</td>
<td>1.054</td>
<td>1.656</td>
</tr>
<tr>
<td>Commercial</td>
<td>0.001226</td>
<td>1.738</td>
<td>2.817</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>-0.05311</td>
<td>4.269</td>
<td>4.744</td>
</tr>
</tbody>
</table>
Figure Captions:

Figure 1. Empirical Semivariogram clouds and Isotropic fitting model. Note: distance h in foot.

Figure 2. Semivariogram cloud in the direction of NNW. Note: distance h in foot.

Figure 3. Semivariogram cloud in the direction of WSW. Note: distance h in foot.

Figure 4. Searching neighborhood for the overall model.

Figure 5. Land value surface in Milwaukee.

Figure 6. Land value surface TIN in Milwaukee.

Figure 7. Residential land value surface TIN.

Figure 8. Manufacturing land value surface TIN.

Figure 9. Commercial land value surface TIN.

Figure 10. Distribution of African-American population in the City of Milwaukee, 2000.
Figure 1.
Figure 2.
Figure 3.
Figure 5
Figure 9.
Figure 10.