GIS SUPPORTED HEDONIC MODEL FOR ASSESSING PROPERTY VALUE IN WEST OAKLAND, CALIFORNIA

ZHOU De-min^{1,3,4}, XU Jian-chun², GONG Hui-li³

- Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130012, P. R. China;
 Geographical Information Science Center, University of California, Berkeley, CA 94720, U. S. A.;
 Key Laboratory
 - of Resource, Environment and GIS, Capital Normal University, Beijing 100037, P. R. China;
 - 4. Graduate School of Chinese Academy of Sciences, Beijing 100049, P. R. China)

ABSTRACT: A hedonic linear regression model is constructed in this paper to estimate property value. In our model, the property value (sales price) is a function of several selected variables such as the property characteristics, social neighborhoods, level of neighborhood environmental contaminations, level of neighborhood crimes, and locational accessibility to jobs or services. Definitions and calculation of these variables are approached by using Geographic Information System tools. For improving estimation, gravity model is employed to measure both levels of neighborhood toxic sites and crimes; and a time-based method is used to measure the locational accessibility rather than simple straight-line distance measurement. This study discovers that the relationship between house value and its nearby highway is nonlinear. The methodology could help policy makers assess the external effects of a property. Our model also could be used potentially to identify the current and historic trends of development caused by neighborhood or environments change in the study area.

KEY WORDS: GIS; property value; neighborhood effect; Hedonic Price Analysis

CLC number: P208 Document code: A Article ID: 1002-0063(2006)01-0070-09

1 INTRODUCTION

Oakland, located at 37 47 43 N, 122 43 41 W, is a city on the east side of San Francisco Bay in Northern California in the United States. As of 2000, the city had an area of 202.4km² with a total population of 399 484, making it the third largest city in the San Francisco Bay Area after San Jose and San Francisco. The city of Oakland stretches from San Francisco Bay up into the East Bay Hills (from the "flatlands" West Oakland, North Oakland up into the foothill districts East Oakland). In 2000, the population density was 2751.4/km², and racial makeup of the city was 35.66% African American, 31.29% White, 15.23% Asian, and 17.82% from other races (US Census Bureau, 2000). West Oakland, home of the Port of Oakland, is one of three major shipping ports on the US West coast. So West Oakland is regarded as historically working-class areas. Recently, however, the character of west Oakland began to change into more technical and skillful industries as waves and waves of immigrants coming from both within the US and other countries, which have led real estate prices to skyrocket in the past decade. Regional planning strategy encourages redeveloping those toxic sites in West Oakland. Before the clean-up done as part of the redevelopment process, it is necessary to identify the potential of currently toxic places for future development, and furthermore, reassess their value for specific redevelopment objectives.

Assessment and prediction of property value has being focused since a huge potential benefit always comes with functional change of land use, especially for those unused or environmental contamination sites in cities. It is well known that redevelopment in a city regularly causes revaluation of properties, and decrease or increase of the property is largely depended on how property and its environmental characteristics to be changed. LI and BROWN (1980) tested the impacts of three types of micro neighborhood variables (aesthetic attributes, pollu-

Received date: 2005-10-30

Foundation item: Under the auspices of the Research Client West Oakland Environmental Indicators Taskforce, Talented Foundation of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences (No. C08Y17)

Biography: ZHOU De-min (1967-), male, a native of Dexing of Jiangxi Province, associate professor, specialized in geographic information science and environmental sciences. E-mail: zhoudemin@neigae.ac.cn

tion levels, and proximity) on housing values. Their findings suggested that housing prices rose due to accessibility, but fell due to problems such as congestion, pollution or unsightliness. Some of earlier articles have explored the effects of some non-residential uses, such as commercial, office, industry, on neighborhood property value. Few these articles, however, have developed a standard method, incorporated into Geographic Information System, for reassessing residential redevelopment by calculating neighborhood and environmental effects (SONG and KNAAP, 2004).

Hedonic hypothesis is that goods are valued for their utility-bearing attributes. Statistical methods such as regression analysis are used to measure the value of a particular attribute, taking into account additional attributes associated with the particular good under study. The hedonic price function is specified that the residential sales price of a house is a function of neighborhood and environmental characteristics, structural characteristics of the house. Many hedonic studies examine the value of attributes that contribute to overall housing value. In these studies housing prices are often used as the dependent variable and explanatory variables generally include structural characteristics of the house, neighborhood, and measures of environmental quality (KIEL and ZABEL, 2001). In general, hedonic pricing models assume that the housing value is comprised of a bundle of characteristics (GEOGHEGAN, 2002):

$$R = +S + L + G +$$
 (1)

where R is an (n \times 1) vector of housing prices, S is an (n \times k) matrix of parcel/structure characteristics, L is an (n \times 1) matrix of neighborhood characteristics, G is an (n \times 1) matrix of spatial and locational variables, , , are the associated parameter vectors, and is an (n \times 1) vector of random error terms.

In the analysis that follows, we considered four sets of characteristics that affect the housing value: 1) physical housing attributes; 2) location/accessibility; 3) disamenities such as both neighborhood levels of toxic contaminations and crimes; and 4) socio-economic characteristics/neighborhoods (Table 1). All these characteristics, which measured by using GIS methods and the neighborhood demographic profiles of each house sold in the area from 1997 to 2000, finally incorporated into a hedonic price linear regression model to determine what factors cause price differentials between properties there (Fig. 1). Variables that were statistically significant in the determination of property sales prices were identified. The variation of these variables across West Oakland was all thematically mapped.

After a regression equation was set up, it was applied to predicting the value of specific sites where assuming the construction of a new single-family, two-bedroom, one-bath home at each location. The test of the model was run on a group of properties in the southern area of West Oakland, where approximately 20 toxic sites within a quarter mile of these properties were taken out of the data set to represent cleanup activities. The model pre-

Table 1 Description of candidate variable for model

Property	Variable code	Definition
Property value	ADJ_PRICE	Sales price of each MLS (Multiple Listing Service Dataset) property in year 2000 equivalent dollars
Physical hous- ing attribute	GARAGE POOL BLD_AGE BR BA LOT_SQFT SQFT	Presence of a garage at the MLS property Presence of a pool at the MLS property Age of the residential building on the MLS property Number of bedrooms in the building on the MLS property Number of bathrooms in the building on the MLS property Total square feet of each MLS property lot Total square feet of floor space of each MLS property residential building
Location/accessibility	BARTWALK BARTCAR DT_COST FWYCAR	Dummy variable for MLS properties in 5 min. walking distance to a BART (Bay Area Rapid Transit System in San Francisco) station Travel time by car to nearest BART station from MLS properties Travel time by car to Downtown Oakland from MLS properties Travel time by car to nearest freeway on ramp from MLS properties
Social demo- graphics/neigh- borhoods	PER_WHT90 INCOME_PE DENSITY	Percentage of persons that were the White in 1990 in the census block of each MLS property Per capital income in 1990 of persons in the census block of each MLS Number of persons per acre in the census block of each MLS property
Environmental hazard	TOX DISTANCE	Number of toxic contamination sites within 1/4 miles of each MLS property, weighted by distance from property location Shortest straight-line distance to a freeway structure from MLS properties
Surround disamenity	CRIME	Number of crimes reported for 1999-2000 within 1/4 mile of each MLS property location, weighted by distance from property location

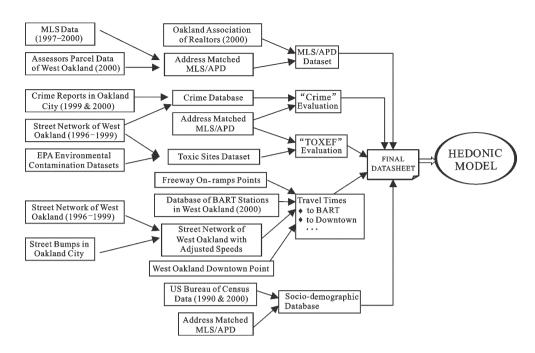


Fig. 1 Overview of GIS supported hedonic model for assessing property value in West Oakland

dicted that the 127 properties in this area would increase in value by an average of US\$14 000, and all these recyclable properties would be roughly US\$1.8 million. Our model demonstrated that the housing value was not only determined by the housing physical features such as house age, house square footage, but also largely affected by proximity of highway, neighborhood environmental contaminations and neighborhood crimes.

2 METHOD

2.1 Variables Selection and Data Collection In explaining the variation in the transaction price of the parcel, four types of explanatory variables were used, most of which were the specific variables for the general variables in Equation (1). All definitions of the candidate variables for regression model were briefly described in Table 1.

The sales price of each property were served as the dependent variable, adjusted by the average rate of real estate inflation for the area, and normalized to a base year. Two datasets were available for our research, each has its advantage: Multiple Listing Service (MLS) data provides an accurate report of the market value of each property listed, however, some addresses of property missed. Assessor's Parcel Data (APD) of West Oakland offers a larger pool of data points and precise address. So combination of two datasets made us to get satisfactory data in the study area. All residential property sale prices

from 1997 to 2000 were compiled.

Since property sales prices vary according to the attributes of the land and improvements therein, each property shall be described using a set of variables that captures these attributes (GOODMAN and THI-BODEAU, 2003). Wide ranges of these variables were observed, finally, total building area, number of bedrooms, number of bathrooms, and age of structure were selected as candidate variables for assessing model. These data could also be acquired from the MLS dataset. The accessibility of each property in relation to major employment centers can be a critical factor that determines its desirability. Most previous studies employed the shortest distance to Central Business Districts (CBD) or major commercial activity centers to measure accessibility to commercial uses (SONG and KNAAP, 2003; GEOGHEGAN et al., 1997). However, it is a daunting task to identify each employment center in the region, and measure the relative accessibility of each property to each employment center. Therefore, we proposed to identify the major ingress and egress transportation points to West Oakland that connect it to the rest of the Bay Area. Accessibility could be measured by the travel time using various models to reach each of these transportation nodes. Although travel times cannot be directly measured with available data, we can calculate the relative accessibility of properties to jobs and the transportation network. Finally, variables are defined by: travel time to freeway on-ramps by car, travel time to BART

stations by car, travel time to downtown Oakland by car, travel time to BART by walking.

The type of neighborhood can make a huge difference in a property market value. Unfortunately, the racial or economic composition of neighborhoods may be an important factor (LI and BROWN, 1980). This is particularly true in West Oakland, an area being traditionally viewed as a predominantly African American and poor area. More than 20% of the population was below the poverty line in 2000 (US Census Bureau, 2000). These factors have likely been a hindrance to investors in the past. To account for social neighborhood characteristics, the third set of variables were defined by: Percent White Persons in 2000 by Census Tract; Percent White Persons in 1990 by Census Tract; Change in Percent White from 1990 to 2000 (Fig. 2); Average Household Income in 2000; Percent of Residents below the Poverty Line in 2000.

The presence of environmental hazards, including soil pollution and toxic industrial emissions, can play an important role in the determination of a neighborhood's desirability and performance in the real estate market (JACKSON, 2001; DEATON and HOEHN, 2004). West Oakland confronts a legacy of groundwater and soil contamination after decades of use as an industrial center for the East Bay and a transportation node for commercial activities. This study proposed to measure the proximity of properties listed in the MLS dataset to those environmental hazard sites identified by the EPA (and any other available source). The variable, level of neighborhood environmental hazards, was judged by proximity of each property to toxic sites.

The level of neighborhood crime can also significantly affect the sales price of property. Oakland's bad reputation came with a high-crime rate in the city. Our model used Oakland Police Crime Reports to identify locations

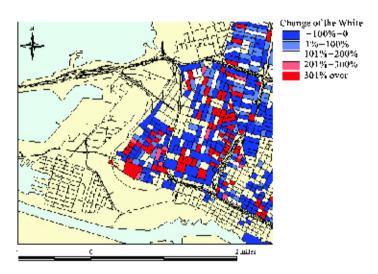


Fig. 2 Map of change in percent White from 1990 to 2000

of crimes. The number of crimes for each study year (1997-2000) was compiled and geocoded. Each census block group then was classified according to the level of crime that had been recorded there. Once done, then each property's neighborhood crime level was included in the regression model as an additional attribute field for each property.

2.2 Data Processing

2.2.1 Accessibility Measure

We hypothesized that the main employment centers for West Oakland are located within downtown Oakland. We further assumed that car or BART would be the primary modes of travel to reach these jobs. Local street network was used for accessing the freeway network and BART systems from West Oakland neighborhoods.

While a simple measure of the straight-line distance from each property to Central Business Districts (CBD) or major commercial activity centers would be a too simple measure to capture the delays associated with driving on congested or low speed limit streets as well as indirect routes navigating a grid street pattern as seen in West Oakland. Consequently, we focused on estimating the travel time necessary on each street link, as determined by factors such as posted speeds and the presence of traffic calming devices.

1996 TIGER/Line streets file was applied to estimating the posted speed limits of each link in the West Oakland network. Each link within the network comes with an associated speed limit (usually 55m.p.h. for freeways, 20 to 30m.p.h. for common roads, and 25 or 35m.p.h. for access ramps). The data of hump number and its location

were obtained from a spreadsheet file that listed the locations of all speed bumps in West Oakland by the Oakland Traffic Department; we assumed that the speed along those links would be decreased by 4m.p.h. per bump from the speed limit estimated in the TIGER/Line streets file. Each link's speed has been adjusted as follows:

Adjusted Speed of Link = Given Speed of Link - (4 * # of Bumps on Link)

Then, a coefficient called impedance was calculated from this new speed. "Impedance" was expressed in time units (in minutes). We defined impedance for each link by:

Link Impedance=Length of Link/Adjusted Speed of Link

The Network Analysis in ARCView was used to find and display the areas accessible from each of our 14 accessibility points by the amount of travel time (the relative impedance). For each accessibility point (freeway on-ramp, BART station, or downtown), we directed ARCView to draw a series of polygon buffers around the points, representing the amount of time it would take to drive from within each band buffer to the point of attraction. As a result, we were able to display the automobile accessibility zones within West Oakland.

To measure the four BART stations within West Oakland, we assumed that people were walking at a speed of

3m.p.h., which enables them to cover 1/4 miles in five minutes—a standard metric of walking accessibility for neighborhood transit stations. As opposed to the automobile travel, which can cover great distances with relative ease, we assumed that the number of potential transit riders accessing a station by walking tends to drop of rapidly once you get beyond this five-minute walking distance. Therefore, in this case, we coded each MLS property data point with a "1", representing that a property is within this five-minute buffer, or "0", indicating it is outside the five-minute buffer. The relative accessibility buffer polygons for each station area were calculated in ARCView using the same impedance formulas as used for automobiles, except the automobile speeds were replaced with walking speeds (3m.p.h.) and there was no impedance associated with speed bumps.

Using these methods, the accessibility (travel time impedance) value of each property to each of the 14 accessibility points was stored as four separate variables leading to four accessibility maps (Fig. 3): BART by car accessibility, BART by walking accessibility, freeway on-ramps by car accessibility, and accessibility to downtown by car. By overlaying these maps on the MLS, Vacant, and Sub-vacant geo-coded property data points, we were able to classify each property by its relative accessibility.

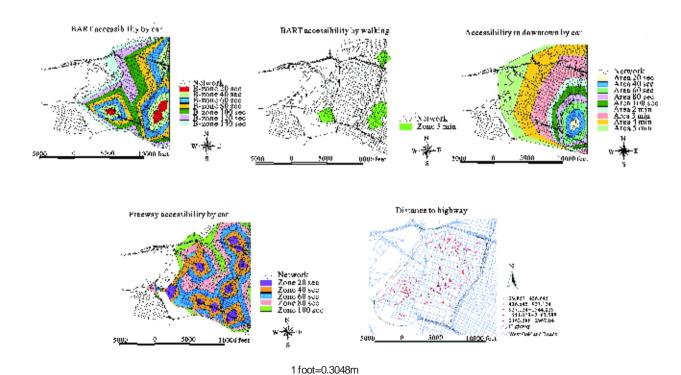


Fig. 3 Maps of accessibility value of properties in study area

2.2.2 Toxic influence measure

Nearby environmental toxic sites could influence land

value of properties, which could be measured by proxmity of each property to toxic sites. It has 3 notable characteristics:

- (1) This effect is constrained by a certain distance upper bound. Property value is not affected by toxic sites beyond this distance limit.
- (2) Inside this distance limit, every toxic site affects the property value and their effects accumulate.
- (3) However, the contribution of each toxic site is different in that they are in different pollution status and their distance to the property varies. Generally the more a site is contaminated and the closer a site is to the property, the larger the effect will be.

This problem could be addressed by gravity model (MIKKONEN and LUOMA, 1999). Gravity model addresses the situation where 2 variables interact with each other spatially, and their interaction decreases by distance. A general form of the gravity model is:

$$W_{i} = \sum_{j=1}^{n} \frac{P_{j}}{d_{ij}^{b}}$$
 (2)

where W_i is the proximity measure for location i to be a member from the same population as j, i is a location in space where we wish to determine the probability of finding a member from the same population, P_j is the population size at j, d_{ij} is the distance between i and j, n is the total number of population centers observed j, b is a parameter affecting the weighting of impact of distance.

Based on the 3 constraints above, we apply a simple gravity model to assessing proximity of a property to toxic sites:

$$TOXEFF = \sum_{i=1}^{n} \frac{tox_{i}^{pw}}{dist_{i}^{pd}}$$
 (3)

where TOXEFF is the effect of toxic sites to the value of a property; pw, pd adjust the relative contribution of tox and dist; tox_i measures the degree of contamination of a toxic site i. The more serious the pollution, the bigger the tox; dist_i measures the distance between a property and a toxic site i; n counts the effective toxic sites around a destination property. A toxic site is effective when the dist is less than a certain distance limit.

The distance limit and the parameters pw, pd are relevant to people's notion of toxic sites and distance. In transportation field this distance limit is usually set to the average walking distance in 5 minutes (1/4 miles). tox and pw control the effect on property value of toxic sites from the same distance. In our data set most of the tox were 1, some of them ranged from 1 to 4. We assumed that from the same distance 2 minor toxic sites (e.g., tox=1) had the same or larger effect as 1 major toxic sites (e.g., tox=4, here tox is a very relative measurement), and we selected pw=0.5. We also assumed that with the same tox, 1 toxic site 100 feet away had the same effect as 4 toxic sites 1600 feet away, then we selected pd=0.5.

With pw=pd=0.5, and effective distance of toxic sites to be 1/4 miles, we calculated the toxeff of toxic sites to properties listed in MLS data (Fig. 4). Higher value of toxeff appeared where toxic sites were dense. The model with the above parameters addressed the problem correctly.

2.2.3 Crime influence measure

Similar to toxic sites, crime events have similar effect on property value and the 3 similar constraints on the effect. So a similar gravity model also was applied to addressing the crime effect on property value:

$$CRIME_EFF = \sum_{i=1}^{n} \frac{1}{dist_{i}^{pd}}$$
 (4)

where CRIME EFF is the effect of crimes to the value

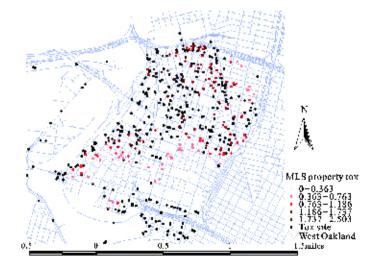


Fig. 4 Map of TOXEFF value of toxic sites to properties in study area

of a property; dist_i measures the distance between a property and a crime location i, pd adjusts the relative contribution dist; n counts the effective crimes around a destination property. A crime is effective when the dist is less than a certain distance limit.

Similar to modeling environmental toxic sites, we selected 1/4 miles as the distance limit. Again, the pd is relevant to people's notion of distance of crime. From discussion in our group we assumed 1 crime event 100 feet away had similar effect to 12 crimes reported 1600 feet away, and we selected pd=0.9. With these parameters, we ran the above model and calculated the CRIME_EFF of crime events reported to properties listed in MLS data set, and showed it in a map (Fig. 5). Higher value of CRIME_EFF appeared where the crime events were dense. Our model with above parameters correctly measured the crime influence on property.

2.3 Modeling

Once the final datasheet file was created containing all the variables available that conceivably might have an effect on residential property values in West Oakland, this file was imported into a statistical package capable of running multinomial linear regressions (S-Plus and SPSS were used). A series of exploratory tests were run to determine which variables showed a statistically significant relationship with residential property sales prices in West Oakland. First, the full compliment of variables was inserted into a "stepwise" linear regression equation for an attempt to find the combination of variables with the "best fit." We hoped to find a set of variables that would explain the maximum amount of variation in the residential property values of West Oakland. Next, a series of Pearson's bivariate correlation coefficients were run to identify those variables that were highly correlated with

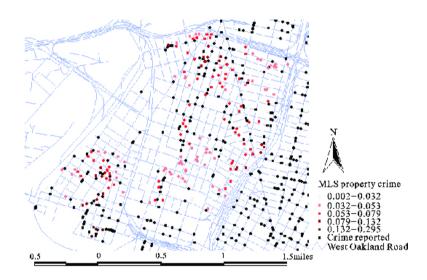


Fig. 5 Map of CRIME_EFF value of a property in study area

each other and might cause multicollinearity.

Once the best fitting model was identified and multicollinearity was minimized, a final linear regression model with the select group of independent variables was run. Results for this model run were shown in Eq. (5), and predicted property value from this model was shown in Fig. 6. The variable components of the final model were all statistically significant at the p < 0.05 level, the R-Square result for this final model was 0.595, indicating that the variables we selected described roughly sixty percent of the variation in the sales price data.

$$y=94523.03-346.32x_1+7135.18x_2+26324.22x_3-37970.5x_4-304756.97x_5-17048.08x_6+34.11x_7+26.25x_8 \eqno(5)$$
 where y represented the sales prices; x_1 represented the

age of a residential building; x_2 was the number of bedrooms in the building; x_3 was the number of bathrooms; x_4 was travel time by car to nearest BART station; x_5 was number of crimes; x_6 was number of toxic contamination sites weighted by distance; x_7 was the shortest straight-line distance to a freeway structure; x_8 was total square feet of floor space of each building.

3 RESULTS AND DISCUSSION

The regression results demonstrated that some of the coefficients made immediate sense. For example, the age of the house is negatively associated with house prices: the older the building, the lower the sales price will be. The same intuition applies to the impact of neighbor-

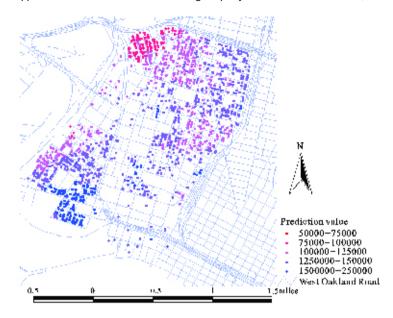


Fig. 6 Map of predicted value of property in study area

hood toxic contamination and crime: the closer a home is to a single or cluster of contaminated site or a site with more crimes, the lower the perceived quality of life will be in the area and sales prices will be lower as well to reflect this. Similarly, there is a positive correlation between a house price with number of bathrooms, bedrooms, and square footage: the more rooms and the bigger the house is, the more it is worth. The research results are consistent with some previous studies (KIEL and ZABEL, 2001; SONG and KNAAP, 2003).

For testing that neighborhood highway as a complex variable factor, being both amenity of traffic accessibility and disamenity of environmental pollution, two variables of FWYCAR and DISTANCE (Fig. 3) were defined as candidate variables (Table 1). The former measured the amenity effect of the traffic accessibility of property, while the later measured the disamenity effect of the proximity of property to the nearest a freeway structure. The regression model showed that property sales prices had a positive association with DISTANCE and a negative association with FWYCAR, which means that the longer it takes to drive to get to a freeway on ramp, the less the house is worth, as well as the farther away from a freeway structure, the more the house is worth. A reasonable interpretation for the results is that proximity of freeway structures associated noise, air pollution, and traffic depresses housing values since they tend to negatively affect quality of life (BOYLE and KIEL, 2001). However, residential property values benefit from traffic accessibility to a freeway on ramp, allowing quick access to regional services and employment. Thus, we believed that an optimal

residential location would be far away from a freeway structure to avoid the negative externalities of these facilities, but within a short driving time of a freeway on ramp to enjoy the benefits of regional mobility. The result indicated that the relationship between housing value and nearby highway is nonlinear, which demonstrated our consumption of neighborhood highway to be both disamenity and amenity.

4 CONCLUSION

In this paper, we focused on West Oakland where some toxic sites and surround vacant properties would be encouraged for redevelopment as part of the required planning and development process in the city. Four sets of quantitative measures, physical housing attributes, location/accessibility, socio-economic characteristics/neighborhoods, and both levels of surround crimes and toxic contaminations, were approached by GIS methods and then incorporated with a hedonic regression analysis to estimate property value. Our research demonstrated that the most predictable variables included the age of the home, neighborhood toxic contamination, neighborhood crimes, number of house bathrooms, house bedrooms, and house square footage, the proximity of properties to the nearest freeway structure and the drive time from the house to a freeway on ramp. A test of the model predicted that the 127 properties in this area would increase in value by an average of US\$14 000 with toxics cleanup activities, and all these recyclable properties would be roughly US\$1.8 × 10^{6} .

ACKNOWLEDGMENTS

I am grateful to Prof. John Radke for providing me with consistent encouragements and valuable suggestions. Many thanks for project group team member Ferrell Christopher, Julie Raffaillac, Rene F Poitevin and Shinichiro Ikeda. This paper was conducted during the first author's academic visiting in the Geographical Information Science Center, University of California at Berkeley.

REFERENCES

- BOYLE M A, KIEL K A, 2001. A survey of house price hedonic studies of the impact of environmental externalities [J]. J. Real Estate Literate, 9: 117-144.
- DEATON B J, HOEHN P J, 2004. Hedonic analysis of hazardous waste sites in the presence of other urban disamenities [J]. Environmental Science & Policy, 7(6): 499-508.
- GEOGHEGAN J, 2002. The value of open spaces in residential land use [J]. Land Use Policy, 19(1): 91-98.
- GEOGHEGAN J, WAINGER L A, BOCKSTAEL N E, 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS [J]. Ecological Economics,

- 23: 251- 264.
- GOODMAN C A, THIBODEAU G T, 2003. Housing market segmentation and hedonic prediction accuracy [J]. Journal of Housing Economics, 12 (3): 181-201.
- JACKSON T O, 2001. The effects of environmental contamination on real estate: a literature review [J]. J. Real Estate Literate, 9: 93-114.
- KIEL K A, ZABEL J, 2001. Estimating the economic benefits of cleaning up Superfund sites: the case of Woburn, Massachusetts [J]. J. Real Estate Finance Econ., 2: 163-184.
- LI M M, BROWN H J, 1980. Micro-neighborhood externalities and hedonic housing prices [J]. Land Economics, 56(2): 125-141.
- MIKKONEN K, LUOMA M, 1999. The parameters of the gravity model are changing-how and why?[J]. Journal of Transport Geography, 7(4): 277-283.
- SONG Y, KNAAPG J, 2003. New urbanism and housing values: a disaggregate assessment [J]. Journal of Urban Economics, 54: 218-238.
- SONG Y, KKNAAP G J, 2004. Measuring the effects of mixed land uses on housing values [J]. Regional Science and Urban Economics, 34: 663-680.
- US Census Bureau, 2000. Profile of general demographic characteristics: city of Oakland, California [EB/OL]. Available from: http://factfinder.census.gov/servlet/SAFFFacts