ANTICIPATING LAND USE CHANGE USING GEOGRAPHICALLY WEIGHTED REGRESSION MODELS FOR DISCRETE RESPONSE

Yiyi Wang
Graduate Student Researcher
The University of Texas at Austin
6.510. Cockrell Jr. Hall
Austin, TX 78712-1076
yiyiwang@mail.utexas.edu

Kara M. Kockelman
(Corresponding author)
Professor and William J. Murray Jr. Fellow
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076
kkockelm@mail.utexas.edu
Phone: 512-471-0210
FAX: 512-475-8744

Xiaokun (Cara) Wang
Assistant Professor
Department of Civil and Environmental Engineering
Bucknell University, Lewisburg PA through 12/2010
& Rensselaer Polytechnic Institute, Troy NY after 1/2011
cara.wang@bucknell.edu
Phone: 570- 577-1112

The following paper is a pre-print and the final publication can be found in
Transportation Research Record No. 2245: 111-123, 2011
Presented at the 90th Annual Meeting of the Transportation Research Board, January 2011

ABSTRACT

Geographically weighted regression (GWR) enjoys wide application in regional science, thanks to its relatively straightforward formulation and explicit treatment of spatial effects. However, its application to discrete-response data sets and land use change at the level of urban parcels has remained a novelty. This work combines logit specifications with GWR techniques to anticipate five categories of land use change in Austin, Texas while controlling for parcel geometry, slope, regional accessibility, local population density, and distances to Austin's downtown and various roadway types.

Results of this multinomial logit GWR model suggest spatial variations in – and significant influence – of these covariates, especially roadway vicinity and regional access. For example, a one-percent increase in the distance on an undeveloped parcel's distance to its nearest freeway is estimated, on average, to increase the probability of residential development by 1.2%, while the same increase in distance to a major arterial is estimated to increase the probability by 1.8%. Conversely, proximity of roads (via reductions in such distances) is

estimated to boost the likelihood of non-residential development (9.0% in the case of commercial development, for simply a 1% *de*crease in distance to such arterials). The logsum accessibility index is estimated to exert an average positive influence on commercial, office and industrial development tendencies, while dampening land use transitions from an undeveloped state to residential uses. Comparisons of results with a spatial autoregressive binary probit (using all developed land use categories as a single response) and GWR binary probit also provide some insights, with the latter seeming to surpass the former in accounting for spatial effects, as reflected by a lower AIC value.

Key words: multinomial logit, geographically weighted regression, spatial autoregressive probit, land use change modeling

INTRODUCTION

The arena of land use modeling enjoys a variety of approaches. One approach can be found in agent-based models (ABMs), which aim to capture the intrinsic nature of land use patterns by simulating agent-environment interactions (see, e.g., Manson 2000, Parker 2008, Millington et al. 2008, and Zhou and Kockelman 2010). Although computational advances facilitate ABM implementation for complex regions with thousands or more agents, implementation complexity remains a major challenge, along with the lack of formal theories to characterize most agent-environment interactions (e.g., land development decisions) (Parker et al. 2001).

Models of discrete choice are now common in land use modeling. Examples include Verburg et al.'s (2004) series of binomial logit models for residential, industrial/commercial, and recreational land uses on a 500m by 500m grid-cell map, Zhou and Kockelman's (2008) logit models for parcel subdivision, and UrbanSim's (Waddell et al. 2003) rather popular land use modeling code. Even after controlling for a host of local, neighborhood attributes around grid cells and parcels, much spatial autocorrelation can remain in unobserved factors. Very few existing studies attempt to account for such effects, since these imply two-dimensional dependence across, potentially, thousands of observations, requiring manipulation of large matrices and high dimensional multivariate distributions (Wang and Kockelman 2009a, LeSage and Pace 2009).

As with various other socio-economic factors (including home prices, poverty levels, travel distances, and election outcomes), land use patterns tend to be correlated across space. The underlying idea emerges from Tobler's First Law (Tobler 1970): everything is related to everything else, but near things are more related than distant things. Wang and Kockelman (2007) described the three main approaches to address spatial effects: geographically weighted regression (GWR), spatial filtering, and direct incorporation of spatial effects. Spatial filtering removes residual spatial relationships by eliminating correlated data points in the sample. Direct spatial specifications tend to rely on spatial autoregressive (SAR) and spatial moving average (SMA) processes, as described in Lichstein (2003), Anselin (2004), and LeSage and Pace (2009). Recent work on discrete states of land use change with such specifications can be found in Chakir and Parent's (2009) spatial multinomial probit model, Munroe et al.'s (2002) series of binary probit and random-effect probit models using panel techniques, and Wang and Kockelman's (2009a, 2009b, 2009c) spatially ordered probit model with temporal component.

This paper combines discrete choice models with GWR techniques to analyze the influence of various factors on land development in Central Texas' Travis County, over a 5-year period. Although the GWR technique has been used to study limited dependent variables (e.g., crash counts [Hadayeghi et al. 2010] and binary response [Páez 2006]), its application to parcel-level land use modeling is quite new. Thus, this work seeks to contribute to the literature and urban systems forecasting by applying a multinomial logit GWR (MNL GWR) model in the context of parcel-level land development. A SAR probit binary model is also specified, for comparison of parameter estimates and predictive fit – relative to a GWR probit binary model (after collapsing all developed land use into a single category, to enable model prediction). The following sections describe existing work, data sets used, model specifications, and results, and then provide conclusions, as well as suggestions for future study.

LITERATURE REVIEW

A key advantage of GWR is its explicit allowance for local spatial effects in relatively standard regression models (Fotheringham 2003). Its flexible specification also allows one to examine the stability of parameter estimates over space, and thus highlights the robustness (or lack thereof) of the model's structure. In contrast, SAR models and other direct specifications impose added burden on the specification to be "right", since only one equation governs anywhere (though latent spatial effects are permitted to vary) (McMillen 2010).

GWR enjoys broad application, in fields as diverse as ecology, wealth and epidemics (see, e.g., Platt 2004, Ognev-Himmelberger et al. 2009, Atkinson et al. 2003, and Nagaya et al. 2010). Transportation research applications currently exist for traffic count and crash count prediction (Zhao and Park 2004 and Hadayeghi et al. 2010, respectively) across networks. By contrast, GWR's application to land use change at the level of whole parcels and/or for discrete response in urban contexts remains very rare, and so is the subject of interest here.

In the context of land use attributes, Ghosh et al. (2008) analyzed impervious cover proportion via a continuous-response GWR framework, for data points across Minnesota's Twin Cities metro area. Páez (2006) provided a binary-response application, using a binomial probit GWR with heteroscedastic error terms to analyze development of 324 vacant 1-hectare grid cells near California's Bay Area Rapid Transit lines. Between 1965 and 1990 just 61 of the 324 locations developed; and, as expected, the locally estimated GWR models yielded a higher log-likelihood value, than the standard binomial probit model (with spatially invariant parameters).

Luo and Kanala (2008) and McMillan and McDonald (1999) extended GWR to multinomial cases. The former analyzed four types of conversion (from barren, crop/grassland, forest and water uses to urban land use) using a MNL GWR model (in reverse time, since all outcomes are "urban land use" in the end year). The study was based on satellite data for 30 m by 30 m grid cells in Springfield, Missouri, and did not sub-classify urban uses, since satellite images really cannot distinguish rooftops and parking lots into use types (e.g., office versus commercial). McMillen and McDonald (1999) specified an MNL GWR model to analyze the impact of transportation access on Chicago's land use mixing in the 1920s. Their data set was composed of 1,160 blocks, approximately drawn at a 4-block interval, forming a lattice. The response variable (land use type) was categorized as follows: all residential, all commercial, all manufacturing, residential-commercial mix, and residential-commercial-manufacturing mix. In addition to distance metrics to Chicago's central business district (CBD) and Lake Michigan, other access variables included distances to major streets, commuter train stops, freight rail lines and canals and rivers. They concluded that higher access to transport facilities was significantly associated with more mixed-use conditions.

Parcel-level MNL GWR models remain conspicuously absent in the literature, and that is where this work most contributes, along with a comparative look at a SAR model's results (at the binary-outcome level, since SAR MNP [LeSage and Pace 2009] cannot yet handle large sample sizes). Two separate base specifications were implemented, the first anticipating physical changes in parcel shapes over the five-year data window (i.e., parcel merges, subdivision, and no-change conditions) and the second anticipating land use changes (on unchanged parcels) over the same period.

DATA SETS

In order to apply an MNL GWR model to Austin area data, Travis County parcel details were assembled. These include a three-category physical change response variable, a six-category land use response variable, and the following eight regressors: network distance to the regional CBD (*DistCBD*), Euclidean distances to the nearest minor arterial, major arterial and freeway (*DistMnrArt*, *DistMajArt*, and *DistFwy*), Euclidean distances to the nearest water and golf course (*DistWater* and *DistGolf*), soil slope (*Slope*), a logsum measure of accessibility (*Access*), parcel size in acres (*Area*), and parcel perimeter-to-area ratio (*Perim-to-Area*).

Land Use Types

The County of Travis Central Appraisal District's (TCAD's) data sets were used to define those taxable (privately held) parcels, while City of Austin (CoA) data sets aided in the determination of land use type for non-taxable parcels (which are generally missing from appraisal district data) and in identifying undeveloped parcels in the year 2003. The final land use categories consist of undeveloped, residential (both single- and multiple-family dwelling units), commercial (including retail, entertainment and recreational uses), office, and industrial uses. Parks, greenbelts and preserved land were excluded from analysis because these land types are almost always undevelopable (due to planning restrictions, at least in the short to medium term).

In terms of acreage, among the 674,951 acres of land that correspond to the 299,889 parcels encoded in CoA's year 2003 land use map, 41.9 percent (240,143 parcels) were already developed in year 2003, 3.29 percent (2,647 parcels) were in the form of parks and greenbelts , 9.03 percent (589) were otherwise preserved/protected , and 2.39 percent was covered by water (178 shapes). Although parks/greenbelts and water are, in essence, undevelopable, they provide a form of amenity for other, nearby land uses, and may facilitate developments such as houses and restaurants. For this reason, distances to each parcel's nearest water area and park are controlled for in the spatial models.

Undeveloped parcels are vacant parcels with the potential to develop, and thus exclude parks, preserved land, greenbelts and water; the data set includes a total of 48,445 undeveloped parcels in 2003. Among these, 1,951 had undergone subdivision by year 2008, 3,905 had merged into larger parcels by 2003, and the remaining 42,589 experienced no physical changes[1]. Of course, land use development may take place on portions of changed parcels.

[1] ArcGIS's *Spatial join* function was used to join TCAD 2008 parcel centroids to the CoA's 2003 map of undeveloped parcels. Parcel-merge behavior was detected by a zero value of *join count* (which is the number of centroids that fall within the parcel boundary), and subdivision was determined by a *join count* result greater than 1. (A *Join count* value of 1 indicated parcels experiencing no physical changes.)

Due to the computational intensity associated with the nested structure of such a change (e.g., the nature of the physical change and then the new land use types involved), a nested spatial MNL model is left for future improvement.

In this paper, separate GWR MNL models for the occurrence of physical changes in a parcel and for land use type outcomes were constructed using the parcel-level data snapshot at years 2003 and 2008. The physical change model used the 48,445 undeveloped parcels in 2003 with three outcomes: subdivision, merging and no-change. The land use change model was based on parcels undergoing no physical changes, with five possible outcomes: undeveloped, single-family plus multi-family residential, commercial, office, and industrial uses.

Among the 42,589 parcels that experienced no physical change, 64.8 percent remained undeveloped during the 5-year period. Among those that developed by 2008, the vast majority (98.7%) developed into residential uses, as shown in Table 1. Tables 2 and 3 summarize details on all covariates used in the two models, respectively.

Table 1. Land Use Shares for Physically Unchanged Parcels between 2003 and 2008

Land Use	Code	#Obs.	Share
Undeveloped	0	27584	0.648
Single-Res	1	14446	0.339
Multi-Res	2	59	1.39E-3
Commercial	3	209	4.899E-3
Office	4	127	2.991E-3
Industrial	5	164	3.839E-3
Total		*42589*	*1.00*

Such unbalanced land use distributions (in favor of residential and undeveloped uses by year 2008) can result in a singular or nearly singular Hessian matrix for the model's log-likelihood function in certain neighborhoods (especially in highly residential neighborhoods, far from Austin's mixed-use downtown). To counter the impact of unbalanced response-variable conditions around various parcels, a binary probit GWR model was also estimated, as a complement to the MNL GWR model. In this case, all developed land use categories were combined into one category and served as the developed (y=1) alternative, as opposed to the undeveloped base alternative (y=0), and no inestimable situations arose (due to Hessian singularities present for a variety of parcel settings in the 5-level GWR MNL model). As noted earlier, this binary case was used to allow comparison with results of a binary SAR probit model, as discussed later.

Table 2. Covariates for Prediction of Physical Parcel Changes

	Unit	Min	Max	Mean	Median	StdDev
DistCBD	Mile	0	55.40	19.69	18.26	10.29
DistFwy	Mile	0	21.02	6.429	5.011	5.651
DistMajArt	Mile	0	6.139	0.803	0.483	0.855
DistMnrArt	Mile	0	10.62	1.770	1.474	1.368
DistWater	Mile	0	14.56	3.014	1.520	3.259
DistGolf	Mile	0	14.53	2.581	1.944	2.533
Slope	Percent	0	65.49	7.614	5.654	6.703
Peri-to-Area	1/Feet	4.81E-04	1.127	0.036	0.037	0.024

Area	Acres	1.68E-03	1769	5.108	0.298	32.51

Covariates for Land Use Change Prediction

The five land use types described above serve as response categories for land use change from an undeveloped state in 2003. A variety of attributes or "covariates" are expected to influence the various likelihoods of change, including soil slope and parcel geometry, local population density, distance to the region's CBD, distances to various roadway types, and regional accessibility. Table 3 provides summary statistics for all these variables.

Table 3. Summary Statistics of Regressors

	Unit	Min	Max	Mean	Median	StdDev
DistCBD	Mile	7.000E-01	43.84	19.95	18.55	10.44
DistMnrArt	Mile	3.307E-03	10.62	1.767	1.470	1.350
DistMajArt	Mile	2.354E-04	6.139	0.815	0.490	0.852
DistFwy	Mile	1.169E-02	21.02	6.611	5.414	5.661
PopDensity	Persons per Acre	0.000	149.71	1.136	0.022	5.207
Slope	Percent	0.01	74.60	7.829	5.829	6.874
Access	-	1.201	6.729	5.488	5.684	0.932
Peri-to-Area	1/Feet	7.120E-04	1.060	0.036	0.038	0.019
Area	Acres	3.737E-03	1407	3.329	0.283	21.37

While steeper slopes can be difficult to build upon, they also can be more interesting for views and neighborhood aesthetics. Here, slopes first took the form of a raster layer (at 10 m resolution), as obtained by applying ArcGIS's slope function[2] to the U.S. Geology Survey's National Elevation Dataset. To reduce computational demands, these 10 m grid cells were converted to a 30 m point layer, and the *Slope* attribute was averaged (and then these were averaged for each parcel's spatial extent, to use in the regression models).

Zhou and Kockelman (2008) estimated that parcel size and the ratio between perimeter and area are positively associated with residential development, consistent with the more rectangular shapes commonly observed for residential parcels (e.g., 50 ft x 100 ft). In contrast, parcels with commercial, office and civic uses tend to be more square in shape (Zhou and Kockelman 2008). Thus, parcel areas (measured in acres) and the ratio between each parcel's perimeter and area (measured in inverse feet) serve as two other covariates for land use change prediction.

As noted, local densities of development can also incentivize or deter new development. Here, population per acre in 2005 (evaluated at the TAZ level) serves as the mid-point measure of local population density for the 8-year period. This variable was interpolated for each parcel's census tract using a demographic software's (Social Explorer's) year 2000, 2006 and 2007[3] population counts. Tract-level population densities are defined as population counts per tract divided by the tract's land (not water) area, and then these were assigned to TAZs. Densities for years 2001-2005 and 2008 were estimated using an exponential growth

[2] ArcGIS' slope algorithm searches for (and assigns) the maximum among the eight slope values calculated from the run and rise of each grid cell's center-point elevation versus that of its eight surrounding neighbors.

[3] These three years' population counts are at the tract level. The year 2000's data come from the decennial Census aggregates. The other two sets of counts were estimated by data experts at Social Explorer, who "filled in" data missing in the 2006 and 2007 American Community Surveys using PUMS PUMA aggregates. Then, the changes between 2000 and 2006/2007 PUMAs were allocated to census tracts. Therefore, the end result can as comparable to ACS tract level estimates. (Huang 2010)

assumption, such that $y(t)=y(0)\times(1+i)^t$. In this way, future year values were inferred from a past year's value.

Distance to the region's CBD regularly is a powerful covariate in models of land value and land use (see, e.g., Haider and Miller 2000, Srour et al. 2002, Zhou and Kockelman 2008). Here, this attribute was computed as the shortest-path network distance from each parcel's centroid to the Texas State Capitol, based on Travis County's 2005 coded network. Distances to the nearest freeway, major arterial and minor arterial can also play important roles in a site's viability for development (see, e.g., Iacono et al. [2008] and Zhou and Kockelman [2006]), with access of interest to most developers. Visibility from high-flow facilities is probably of great interest to commercial and office sites, while some sound-dampening and safety-enhancing buffer space is desired for most residential users. Again, Euclidean distances from parcel centroids were used, based on shortest paths on Travis County's 2005 network. In addition, distances to amenities like golf courses and bodies of water (as per the CoA's 2003 maps) were evaluated as covarates, since these are noted at times in the literature as contributing to land use change (see, e.g., Lin et al. 2005). Overall, regional accessibility is also a key component of site attractiveness for a variety of use types (see, e.g., Waddell et al. [2003], Sour et al. [2002], Niemeier [1997], and Sermons and Seredich [2001]). Traditional measures of accessibility, such as travel time, distance and cumulative opportunities, are rather simplistic in nature. Fortunately, the expected-maximum-utility or logsum measure obtained from discrete choice models of destination choice can account for the behavioral nature of such choices (see, e.g., Neimeier [1997]). Here, the *Access* variable is computed as follows:

$$Access_{\mathrm{i}} = \ln\left[\sum_{j=1}^{N}\left(\exp\ \beta_1 dist(i,j) + \beta_2 \ln\left(emp_j\right)\right)\right] \tag{1}$$

where $Access_i$ is the accessibility index for location i (i.e., the traffic analysis zone of the parcel i), $dist(i,j)$ is the shortest-path network distance from each origin TAZ i to each destination TAZ j, emp_j denotes total employment in TAZ j, and parameters β_1 and β_2 (-0.226 and +0.269, respectively) were estimated by running a logit model of destination TAZ choice for all 13,942 trips in the 2006 Austin Travel Survey.

METHODOLOGY

This section summarizes the mathematical formulations of MNL and GWR regression techniques, including cross-validation for bandwidth or neighborhood determination (for the spatial weights). Given the site-specific nature of land use data and lack of alternative-specific variables, an unconditional MNL was adopted here, as shown in Eq 2 (and in Greene [2003]).

$$P_{nj} = Prob(Y_n = j|x_n) = \frac{\exp\left(x_n{'}\beta_j\right)}{1+\sum_{k=1}^{J} x_n{'}\beta_k} \tag{2}$$

where n denotes the n^{th} parcel observation, j indexes outcome alternatives (with j=0 indicating the base alternative: a parcel remaining undeveloped in this work's context), and the vector β consists of alternative-specific parameters (to be estimated) for non-generic attributes (such as parcel size and slope). The corresponding log-likelihood is:

$$lnL = \sum_{n=1}^{N}\left[\sum_{j=0}^{K} I_{nj} \ln\left(P_{nj}\right)\right] \tag{3}$$

where I_{nj} is an indicator variable for outcome *j* at parcel *n* and $I_{nj}=1$ if parcel *n* is of land use category *j* and 0 otherwise.

GWR is an extension of weighted least squares (WLS) methods, where the weights are spatial in nature (and falling with separation between observations) and a new regression is run at each data point, to allow parameter estimates to vary over space. In the case of a continuous response (e.g., home prices), Equation 4 shows GWR estimator for the i^{th} data point or parcel:

$$\hat{\beta}(x_i, y_i) = (X' W(x_i, y_i)X)^{-1} X' W(x_i, y_i)Y \tag{4}$$

where $\hat{\beta}(x_i, y_i)$ is the vector of estimated parameters at location (x_i, y_i), X is an n by k matrix of covariates, $W(x_i, y_i)$ is an n by n weight matrix, varying by location (as described below), and Y is an n by 1 vector of response values (across all n neighbors).

As noted, in the current context, the response is discrete multinomial. The log-likelihood function used is that applied by McMillen and McDonald (1999), as shown in Equation 5:

$$lnL_i = \sum_{n=1}^{N} [w_{in} \sum_{j=0}^{K} I_{nj} ln\,(P_{nj})] \tag{5}$$

where w_{in} is the weight for the n^{th} data point with respect to the i^{th} regression point (as described below), and I_{nj} is an indicator variable for land use category *j*. P_{nj} is the probability that undeveloped parcel *n* transitions to land use type *j* by 2008 (as shown in Eq 2). MATLAB software was used to repeatedly maximize this loglikelihood (for neighborhood samples around each parcel), using Newton-Raphson techniques (based on first- and second-order derivatives, as described in Greene 2003).

Weights Used

Fotheringham (2003) describes a variety of weight options. Gaussians weights and their bi-square variation are provided in Equations 6 and 7, respectively. These consider point proximity as well as bandwidth distance.

$$w_{ij} = exp\,[-0.5 \cdot (d_{ij}/b)^2] \tag{6}$$

$$w_{ij} = \begin{cases} [1-(d_{ij}/b)^2]^2, if\ d_{ij} < b \\ 0 \qquad\qquad\quad , otherwise \end{cases} \tag{7}$$

where *b* is bandwidth and d_{ij} is the distance between regression point *i* and data point *j*.

Such functions are called "fixed spatial kernels", indicating that the sample size used for regression at any data point (parcel) *i* is solely determined by bandwidth distance. As noted, data points typically do not scatter evenly over space, so a fixed-distance kernel can cause inadequate sample sizes in sparsely data-populated locations. One remedy is to use "spatially varying kernel", which ensures that effective bandwidths "shrink" in areas where data points are densely distributed and "expand" in sparsely populated locations. An example is the tri-cube weight for pairs of points, expressed as follows:

$$w_{ij} = (1-(d_{ij}/d_{\max})^3)^3 \tag{8}$$

where j is one of point i's N nearest neighbors (otherwise w_{ij} equals zero), and $d_{\max}$ is the distance from the N^{th} nearest neighbor to point i.

Cross-validation (CV) determines the optimal bandwidth value or optimal number of nearest neighbors N by minimizing the CV score (Fotheringham 2003). In essence, the CV technique calculates the sum of squared error terms in each regression point's neighborhood, excluding the regression point itself from the model. This pseudo-sum-of-squared-errors is called the CV score, and is expressed as follows, for the MNL case:

$$CV = \sum_{n=1}^{Nobs} \sum_{j=0}^{J} (I_{\neq n,j} - \hat{P}_{\neq n,j}(b))^2 \tag{9}$$

where $I_{\neq n,j}$ is an indicator variable for data points other than n, so $I_{\neq n,j}=1$ if parcel n is of land use type j, and 0 otherwise. Moreover, $\hat{P}_{\neq n,j}$ is the estimated probability for parcel n having land use type j.

The number of valid data points (N_{obs}) over which the CV score is computed can vary as a function of N. For example, only 25% (1,933) of this work's previous experimental dataset (7,591 data points) yield an invertible Hessian under an $N = 300$ nearest-neighbors specification. This percentage increases to 48% and 58% when N rises to 800 and 1,500, respectively. When N increases to 3,000, the valid percentage falls back to 40%[4]. For comparability across cases, Eq 9's CV score should be normalized, by dividing by the number of valid regression points (rather than simply summing over N_{obs} data points), thus providing an average CV score for each regression point. Given computational intensity, the average CV score was only computed for five N values (N = 300, 800, 1000, 1500 and 3000), yielding average CV scores of 8.29, 5.13, 1.75, 3.82 and 1.98. Thus, parameter estimates were obtained by maximizing the log-likelihood function applicable at each undeveloped parcel, with N =1000; and Hessians were used to compute standard errors and t-statistics, with all results presented below.

MODELS AND RESULTS

This section presents results from the two MNL GWR models, a binary probit GWR model, and a SAR binary-probit model. The binary probit models help counter the estimation issues associated with insufficient variation of land use change in neighborhoods around a variety of parcels, as discussed earlier. The SAR binary-probit was analyzed using Bayesian estimation methods (LeSage 1999), as another point of comparison (with constant parameters and a relatively straightforward spatial representation for dependence in unobserved components).

Results of MNL GWR

As noted earlier, Austin's land use changes (from undeveloped status) heavily favor residential uses, resulting in singular Hessians for sample sets surrounding many parcels. Without a Hessian, one cannot quantify uncertainty in parameter estimates. One remedy is to select regression points (i) in the N = 1000 neighborhoods that contain enough land use-change variation and simply ignore those that do not.

[4] Adding such new variety in data may cause these problems because the spatial weights (for far-away points) fall to levels too low to recognize the added information. The study area's neighborhoods of 800 to 1500 parcels may offer a better mix of different land use types. (Including too many observations will overwhelm the sample with residential parcels thus leading to singular Hessian.) This trend is probably specific to the study area's data sets.

Out of the 33,912 datapoints (a 70% sample of the population) in the physical-change model, 27,174 regression points yield invertible Hessians. Working with these 27,174 data points' results, a variety of estimates emerge, as shown in Table 4. A rise in the *DistCBD* variable's value, on average, is associated with a higher probability for parcels to undergo merging or subdivision. This effect appears stronger in the southeastern part of Travis County than in the northwestern part, as shown in Figure 1. In contrast, an increase in the distance to a parcel's nearest highway tends to dampen such probabilities. Proximity to both the nearest water body and golf course tends to increase the likelihood of merge and subdivision outcomes, with water access, in particular, offering very high elasticities. As expected, larger parcel sizes tend to decrease the likelihood of a merge event but increase subdivision tendencies. The *Perim-to-Area* variable exhibits the opposite effect.

Out of the 42,589 datapoints in the land use model, 3,684 yield invertible Hessians, thanks to a fair amount of land use variation in their neighborhoods[5] (i.e., roughly 7 to 10% of parcels shift to non-residential development). Amid these estimator distributions (one set of estimates per data point), one finds that coefficients on the *Constant* terms for the *Residential* alternative have relatively high mean values, among the four developed alternatives. This is consistent with the fact that the majority of 2003-2008's new developments are residential. The *DistCBD* variable tends to have a negative impact on *Office* development, meaning that office uses are more likely in more central locations, ceteris paribus. This is in contrast to *Residential*, *Commercial* and *Industrial* outcomes′ response to the *DistCBD* variable. For example, new residential developments emerged around the southwestern and northern areas of the county, though many high-rise condominium units have emerged in Austin's downtown.

[5] With $N = 1000$ nearest neighbors (for the tri-cube nearest neighbor weight function), average neighborhood radius was 3.45 miles (with a standard deviation of 1.93 miles).

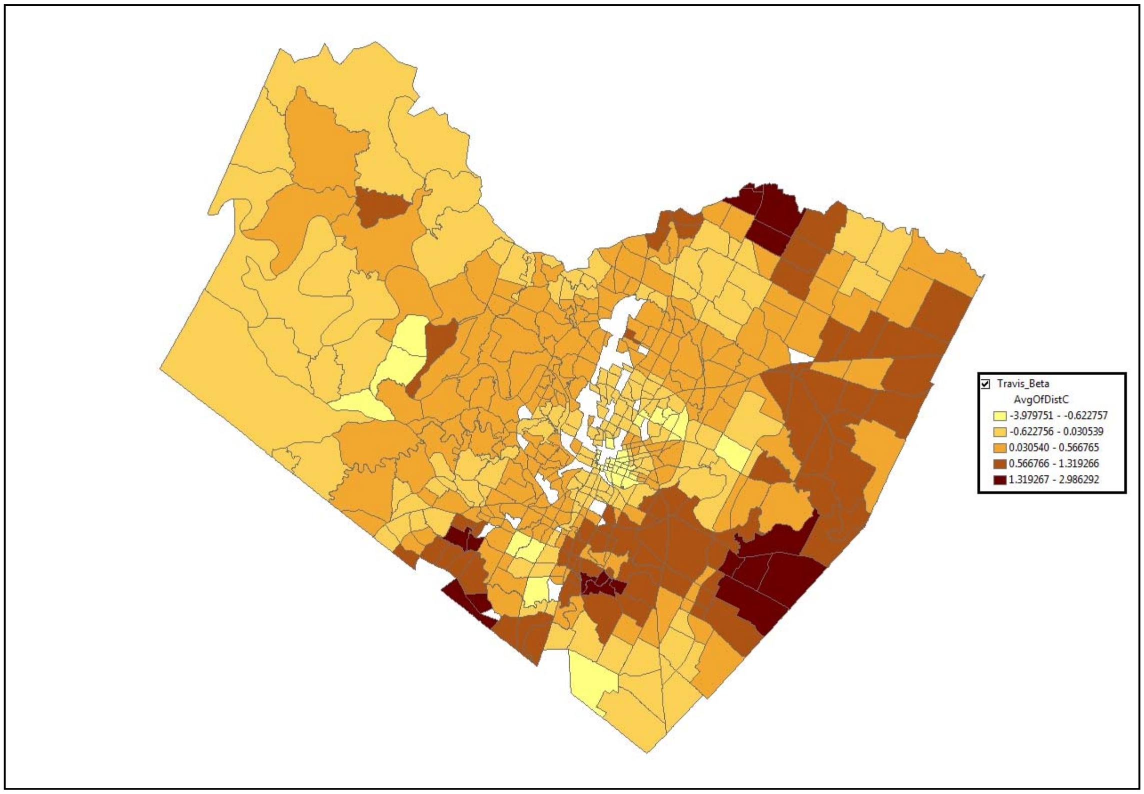

Figure 1. Spatial Distribution of *DistCBD* for Parcel Merge Response

Table 4. Summary Statistics of Parameter Estimates of MNL GWR Model for Physical Change Behavior

			Constant	*DistCBD*	*DistFwy*	*DistMajArt*	*DistMnrArt*	*DistWater*	*DistGolf*	*Area*	*Perim-to-Area*	*Slope*
Merge	Mean	Beta	25.40	0.12	-2.74	-1.87	-2.93	-0.53	0.33	-0.05	21.58	-0.01
		Tstat	-0.67	0.70	-1.05	-1.22	-1.02	-0.36	0.28	0.27	2.00	0.03
		Elasticity	94.42	10.03	-53.51	-45.24	-54.68	-24.78	-0.16	-0.65	68.76	-0.63
	Median	Beta	-1.95	0.06	-0.65	-1.05	-0.64	-0.24	0.15	0.01	19.21	0.00
		Tstat	-0.85	0.42	-0.90	-1.18	-0.98	-0.47	0.16	0.10	2.05	0.13
		Elasticity	-52.66	4.42	-14.66	-23.07	-16.63	-5.44	1.21	0.01	56.67	0.09
	StdDev	Beta	133.83	1.41	10.13	4.25	8.79	6.57	3.84	1.31	35.16	0.12
		Tstat	2.54	1.90	2.17	1.82	2.33	2.02	2.73	0.99	1.78	1.60
		Elasticity	1,804.90	48.88	186.94	95.52	172.17	119.84	62.52	6.48	120.57	3.77
	Min	Beta	-739.86	-25.13	-100.27	-53.72	-94.57	-115.50	-31.92	-26.63	-34.03	-0.96
		Tstat	-6.29	-7.44	-7.71	-7.62	-7.94	-5.42	-9.17	-2.52	-4.27	-4.89
		Elasticity	-151.12	-441.57	-271.55	-182.02	-252.12	-29.39	-34.68	-31.79	-109.43	-35.34
	Max	Beta	134.41	4.53	55.43	22.78	17.65	29.78	97.72	6.56	240.40	0.42
		Tstat	7.46	6.62	4.17	5.77	5.78	6.51	8.33	3.59	6.00	4.17
		Elasticity	2,293.18	276.75	1,221.15	554.93	424.24	506.09 li	1,102.09	1.99	7,175.29	15.84
			Constant	*DistCBD*	*DistFwy*	*DistMajArt*	*DistMnrArt*	*DistWater*	*DistGolf*	*Area*	*Perim-to-Area*	*Slope*
Subdivision	Mean	Beta	40.91	0.19	-3.51	-2.04	-2.87	-1.65	-0.23	0.10	-70.69	-0.05
		Tstat	0.18	0.65	-0.73	-0.90	-0.78	-0.76	-0.20	1.28	-2.52	-0.37
		Elasticity	446.92	2.55	-67.45	-43.75	-70.85	-34.87	-7.31	0.05	-37.54	-2.02
	Median	Beta	2.26	0.10	-0.69	-1.17	-0.76	-0.52	-0.08	0.02	-67.59	-0.03
		Tstat	0.22	0.63	-0.75	-1.19	-0.87	-0.65	-0.13	1.37	-2.57	-0.50
		Elasticity	40.00	6.80	-16.31	-27.42	-23.46	-12.53	-0.99	0.04	-24.08	-0.83
	StdDev	Beta	141.21	1.14	10.48	6.75	9.40	6.98	5.09	0.47	66.62	0.17
		Tstat	1.50	1.75	1.64	1.37	1.31	2.00	1.72	1.02	1.42	1.19
		Elasticity	2,135.53	49.95	29.59	98.46	24.87	33.08	82.17	0.70	28.62	4.62
	Min	Beta	-623.51	-11.40	-18.70	-83.67	-102.52	-35.47	-61.45	-16.92	-51.05	-6.89

		Tstat	-4.22	-4.87	-6.11	-5.71	-5.26	-8.35	-8.08	-3.29	-6.71	-3.34
		Elasticity	-1,368.76	-726.35	-295.58	-128.75	-334.04	-260.17	-131.52	-35.89	-190.13	-20.89
	Max	Beta	1707.71	18.70	52.95	49.23	27.79	30.50	62.17	4.64	39.07	1.01
		Tstat	5.24	7.34	4.98	3.97	4.19	4.62	4.51	3.99	2.11	3.24
		Elasticity	2310.99	454.79	132.94	303.73	493.82	425.60	703.19	2.11	380.56	15.15
Pseudo-R^2_{adj}	Mean	0.987										
	Median	0.987										
	StdDev	0.003										
	Min	0.980										
	Max	0.997										

Table 5. Summary Statistics of Parameter Estimates for the MNL GWR Model of Land Use Change

		Constant	*DistCBD*	*DistMnrArt*	*DistMajArt*	*DistFwy*	*Slope*	*Perim-to-Area*	*Area*	*Access*	*Pop Density*
Residential	Min	-4.889	-0.177	-0.492	-0.028	-0.139	-0.108	-6.788	-0.206	-1.179	-0.132
	Max	6.543	0.287	1.023	2.635	1.216	0.039	13.983	-0.021	0.401	0.013
	Mean	-1.129	0.006	0.128	1.115	0.724	-0.004	3.311	-0.100	-0.071	-0.013
	Median	-0.549	-0.008	0.141	1.176	0.822	-0.003	1.752	-0.105	-0.143	-0.018
	StdDev	1.603	0.098	0.259	0.615	0.273	0.024	3.639	0.030	0.217	6.271E-03
Commercial	Min	-4.439	-0.372	-2.293	-5.013	-1.231	-0.688	-66.143	-0.050	-0.691	-0.021
	Max	4.575	0.243	0.340	-1.764	0.432	-0.381	-0.838	0.014	0.747	0.061
	Mean	-1.784	0.019	-0.832	-3.159	-0.009	-0.499	-12.357	-0.011	0.132	-3.159E-3
	Median	-1.991	0.007	-0.698	-3.038	-0.012	-0.491	-2.210	-0.011	0.195	-5.231E-3
	StdDev	1.403	0.106	0.610	0.547	0.272	0.050	16.433	0.015	0.199	1.652E-3
Office	Min	-9.256	-1.699	-2.714	-6.292	-1.728	-0.291	-159.625	-0.125	-1.863	-0.011
	Max	12.222	0.356	1.579	0.664	1.746	0.052	-19.123	0.046	1.462	0.025
	Mean	-2.701	-0.442	-1.009	-3.395	0.018	-0.102	-62.248	-0.032	0.531	-0.019
	Median	-3.545	-0.465	-1.186	-4.326	-0.102	-0.107	-50.690	-0.021	0.649	-0.021
	StdDev	3.736	0.472	0.775	1.965	0.566	0.056	36.282	0.041	0.670	1.371E-3
Industry	Min	-8.912	-0.250	-3.209	-7.163	-2.102	-0.577	-57.250	-4.048E-	-0.361	-0.522
	Max	1.256	0.419	0.250	-0.225	0.086	-0.359	-5.592	1.118E-02	1.288	0.031
	Mean	-3.438	0.162	-1.598	-2.281	-0.386	-0.474	-18.646	-2.279E-	0.320	-0.073
	Median	-3.699	0.159	-1.756	-2.274	-0.281	-0.483	-15.293	-4.934E-	0.337	-0.054
	StdDev	1.889	0.123	0.933	1.225	0.287	0.054	10.667	9.288E-03	0.279	0.015
Pseudo-R^2_{adj}	Min	0.32									
	Max	0.74									
	Mean	0.63									
	Median	0.57									
	StdDev	0.15									

Table 6. T-Statistics of Parameter Estimates for the MNL GWR Model of Land Use Change

Residential	*DistCBD*	*DistMnrArt*	*DistMajArt*	*DistFwy*	*Slope*	*Perim-to-Area*	*Area*	*Access*	*PopDensity*
Mean	0.306	0.521	4.309	5.353	-0.187	1.161	-4.440	-1.067	-6.721
Median	-0.129	0.715	4.322	5.946	-0.203	1.024	-4.300	-1.995	-5.013
StdDev	1.750	1.456	2.674	1.961	1.040	0.842	0.943	2.104	2.521
Commercial									
Mean	0.192	-1.120	-2.522	0.047	-3.411	-0.864	-0.325	0.371	-0.461
Median	0.034	-1.001	-2.444	-0.044	-3.425	-0.388	-0.410	0.608	-1.232
StdDev	0.657	0.597	0.394	0.617	0.398	0.861	0.629	0.573	0.111
Office									
Mean	-0.911	-0.763	-1.536	-0.100	-0.884	-2.381	-0.385	0.387	-1.163
Median	-1.342	-0.882	-1.851	-0.193	-1.024	-2.260	-0.510	0.789	-1.202
StdDev	1.081	0.601	0.843	0.744	0.495	0.567	1.122	0.951	0.061
Industry									
Mean	0.614	-1.539	-1.285	-0.549	-2.386	-1.056	0.285	0.563	-2.434
Median	0.671	-1.830	-1.371	-0.573	-2.390	-0.910	-0.016	0.727	-2.306
StdDev	0.450	0.700	0.509	0.231	0.597	0.412	0.890	0.489	0.372

Notes: Z=1.96 (5% level, 2-tail); Z= 1.645 (10% level,2-tail).

Table 7. Summary Statistics for Covariates' Elasticity Estimates[6] in MNL GWR Model of Land Use Change

Residential	*DistCBD*	*DistMnrArt*	*DistMajArt*	*DistFwy*	*Slope*	*Perim-to-Area*	*Area*	*Access*	*PopDensity*
Mean	8.823E-03	0.192	1.763	1.224	-0.005	8.013	-4.085E-02	-0.853	-2.321
Median	-0.022	0.244	1.923	1.395	-0.006	4.285	-0.043	-1.772	-2.145
StdDev	0.291	0.407	0.883	0.451	0.037	8.434	0.013	2.557	0.342
Commercial									
Mean	0.088	-2.356	-9.193	-0.030	-1.354	-4.031	-4.784E-03	2.505	-1.598
Median	0.035	-1.979	-8.835	-0.038	-1.328	-8.808	-0.005	3.694	-1.452
StdDev	0.514	1.734	1.595	0.859	0.136	6.178	0.007	3.760	0.532
Office									
Mean	-2.188	-2.870	-9.935	0.061	-0.277	-2.068	-1.422E-02	10.181	-2.397
Median	-2.298	-3.370	-12.657	-0.324	-0.291	-2.672	-0.009	12.407	-2.125
StdDev	2.334	2.204	5.754	1.795	0.153	1.816	0.018	12.791	1.210
Industry									
Mean	0.788	-4.530	-6.651	-1.219	-1.289	-4.526	-1.085E-03	6.057	-3.889
Median	0.783	-4.983	-6.636	-0.884	-1.313	-6.551	0.000	6.413	-5.157
StdDev	0.594	2.642	3.586	0.911	0.146	4.897	0.004	5.289	1.390

[6] Elasticity measures the percentage change in the probability of choosing alternative j that is associated with 1-percentage change in the covariate X_k entering the utility function of that alternative (j), expressed as $E_{j,X_{nj}} = \beta_j \cdot x_{nj} \cdot (1 - P_{nj})$. The elasticity shown for each alternative j is taken as the averaged values of individual elasticities across the various N data points:.$\frac{1}{N} \sum_{n=1}^{N} E_{j,X_{nj}}$.

Distances to each parcel's nearest roadways also exhibit interesting impacts. Holding everything else constant, proximity to major arterials and to freeways appears to significantly suppress new residential development. For display purposes, coefficient estimates on *DistMajArt* for residential and commercial uses were averaged at the TAZ level and are presented in Figures 2 and 3. The northwestern region exhibits a remarkable tendency for residential development to avoid major arterials, with practically significant elasticities throughout. (Note: The magnitudes of these elasticities are in proportion to the sizing of Figure 2's triangle symbols.) By contrast, rising *DistMajArt* (i.e., falling access to major artierals) tends to significantly reduce a parcel's attractiveness for commercial development in the mid-south and northwestern regions of Travis County, as shown in Figure 3. Across the region, a 1-percent increase in *DistFwy* is estimated, on average, to increase the probability of residential development by 1.2%, reflected by an average elasticity of 1.224, and 1-percent increase in *DistMajArt* is estimated to increase that probability by 1.8% (thanks to an average elasticity of 1.763).

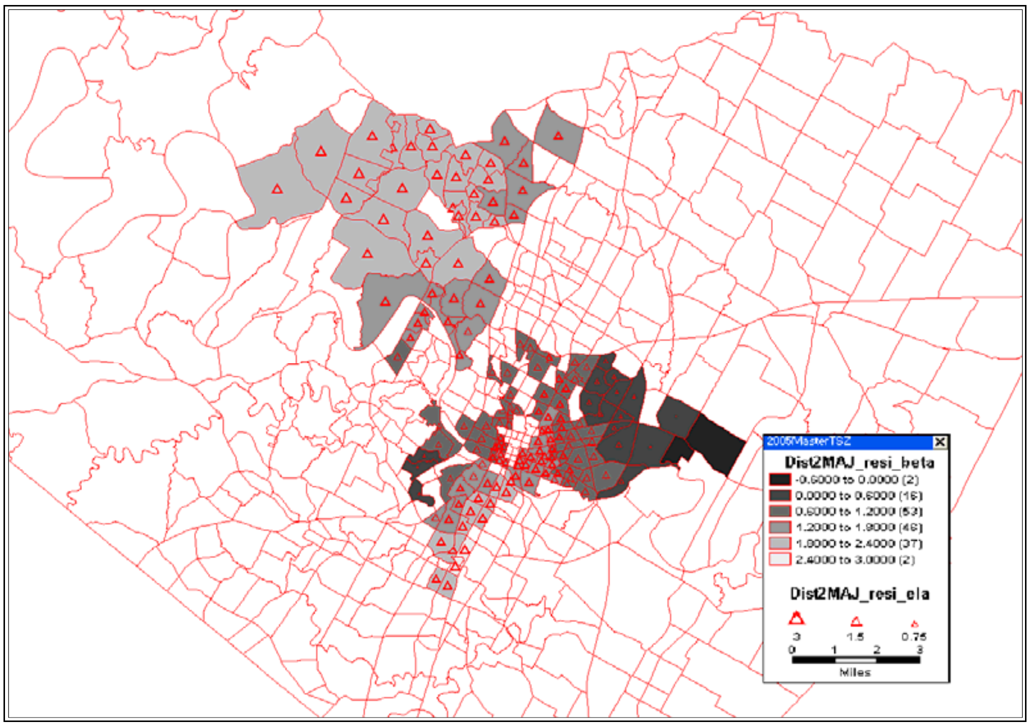

Figure 2. The Impacts of Distance to Major Arterials on Residential Development (Coefficient Estimates and Elasticities across TAZs)

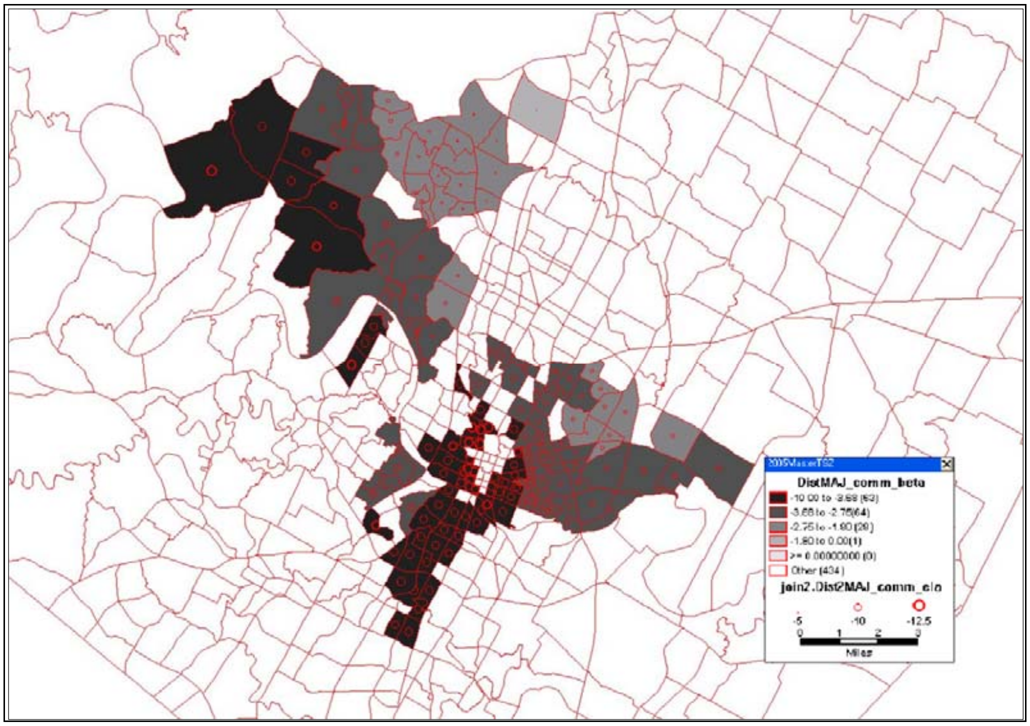

Figure 3. The Impacts of Distance to Major Arterials on Commercial Development (Coefficient Estimates and Elasticities across TAZs)

In contrast, the proximity of roads is estimated to boost the likelihood of non-residential development. For example, a 1-percent increase in *DistMajArt* is associated with a 9.0% decrease in the probability of commercial development. The median elasticity values of *DistFwy* are all negative for commercial, office, service/recreational and industrial uses, suggesting that freeway proximity is meaningful for such non-residential development, ceteris paribus. Similarly, longer distances to major arterials and minor arterials are, on average, inversely associated with commercial, office and industrial development while contributing to residential and service/recreational development[7].

Soil slope tends to have a negative impact on all types of development. Individual elasticities for non-residential developement types are all negative, whereas 1,301 out of 3,684 regression points exhibit positive slope elasticities on residential development (as shown in Figure 4). 120 parcels have elasticity estimates greater than 2.0, indicating that a steeper slope is considered an amenity in these locations, to a practically significant degree. Parcels near Austin's Colorado River tend to have negative slope elasticities because the waterfront region has rather dramatic slopes to begin with, so even a slight increase in this attribute can greatly increase development costs. By comparison, in areas farther away from the waterfront, a moderate increase in slope can offer some scenic benefit, thereby contributing to home development. A *squared Slope* term was also controlled for, but the added correlations resulted in non-singular Hessians for just 435 data points' neighborhood samples.

The influence of the *Perim-to-Area* variable varies across land use alternatives: positive elasticities (averaging +8.0) are estimated for residential development, while average non-

[7] Various recreational land uses (like golf courses and camp grounds) were categorized as the "Service-Rec" use type, which may be causing the positive elasticities of a Service-Rec land use change outcome with respect to the various roadway distance variables.

residential development elasticities are negative. Parcel size tends to be negatively associated with development tendencies, but its influence is not statistically or practically significant (in terms of t-statistics and elasticity estimates).

The logsum *Access* index is estimated, on average, to exert positive influence over commercial, office and industrial development tendencies, while residential development averaged negative t-statistics of -1.995, respectively (with associated median elasticity values of -1.772).

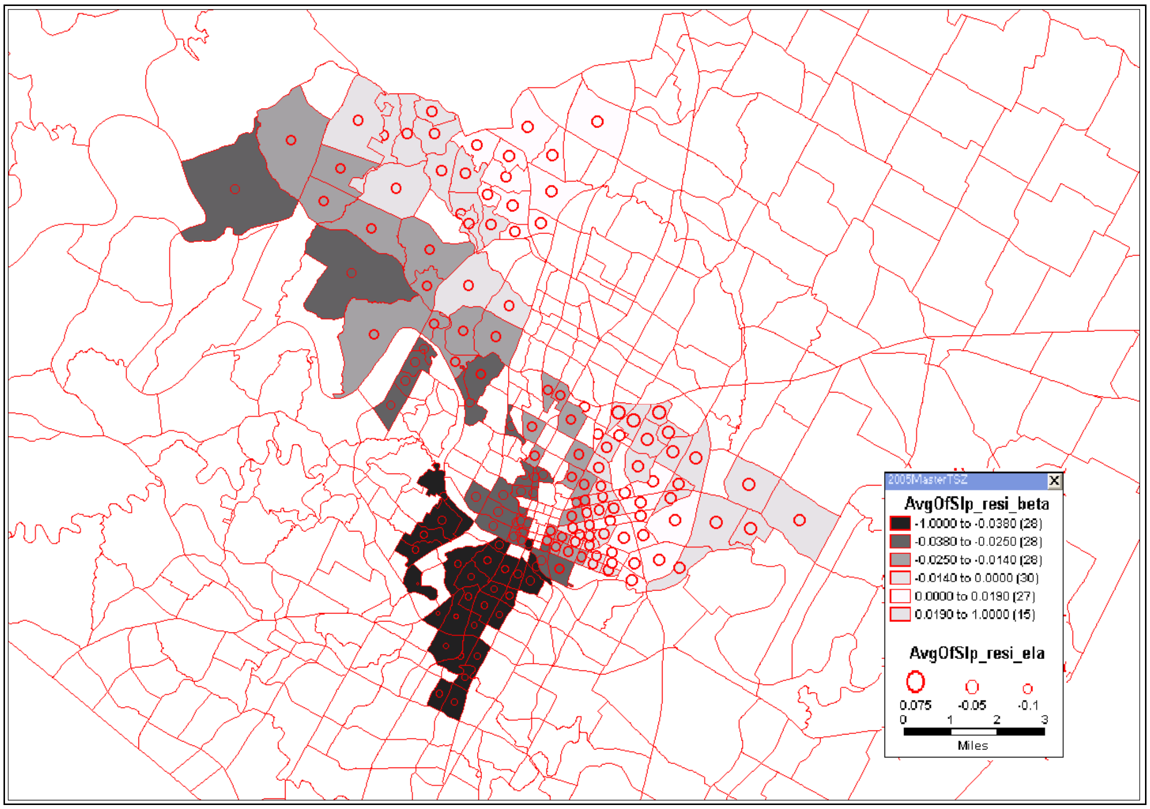

Figure 4. The Impacts of Slope on Residential Development (Coefficient Estimates and Elasticities across TAZs)

Results of the Binary GWR Probit and SAR Probit Models

To avoid the issue of a singular or near-singular Hessian around roughly half the data points in the primary data set, all five developed land uses were collapsed, allowing for a simpler binary specification of land use outcomes in 2008. The likelihood function for a GWR binary probit model is formulated as follows (LeSage 1999):

$$lnL = \sum_{n=1}^{N} w_{in}[(y_n ln\Phi(x_n'\beta_i) + (1 - y_n)\ln{(\Phi(-x_n{'}\beta_i))}] \qquad (10)$$

where $y_n = 1$ (for developed land uses), if $y_n = x_n\beta + \varepsilon_n > 0$, and 0 otherwise. Table 8 summarizes variations in all parameter estimates, their associated t-statistics, and the model's (adjusted) pseudo-R-square values, as run on all 7,951 data points.

Given the fact that the majortiy of developments, if any, are residential in nature, the results of the binary GWR estimates, shown in Table 8, are largely the same as the results for the residential alternative in the MNL GWR case.

Table 8. Summary Statistics of Parameter Estimates, T-Statistics and Adjusted Pseudo-R-Square Values of Binary Probit GWR Model

	Constant	*DistCBD*	*DistMnrArt*	*DistMajArt*	*DistFwy*	*Slope*	*Access*	*Perim-to-Area*	*Area*	Pseudo-R^2_{adj}
Min	-3.89 (-7.48)	-0.29 (-5.88)	-0.84 (-6.29)	-0.81 (-3.10)	-0.27 (-3.51)	-0.14 (-7.58)	-2.99 (-10.6)	-3.79 (-2.47)	-0.033 (-4.62)	0.23
Max	18.75 (9.97)	0.30 (10.55)	0.93 (5.88)	2.93 (10.96)	0.91 (9.60)	0.02 (1.19)	0.20 (2.78)	31.81 (12.11)	0.002 (0.57)	0.25
Mean	1.07 (0.51)	0.02 (1.48)	-0.01 (-0.39)	0.71 (3.88)	0.36 (4.07)	-0.04 (-2.47)	-0.36 (-3.57)	9.18 (4.12)	-0.004 (-1.79)	0.24
Median	0.69 (1.46)	0.04 (1.07)	-0.04 (-0.42)	0.59 (3.66)	0.44 (4.78)	-0.04 (-2.62)	-0.20 (-3.49)	6.74 (3.56)	-0.002 (-1.48)	0.24
StdDev	3.04 (3.72)	0.12 (4.09)	0.20 (2.27)	0.64 (3.06)	0.23 (2.42)	0.04 (1.95)	0.45 (2.68)	9.22 (3.69)	0.004 (1.05)	0.01
Average AIC: 5,100										

Notes: Numbers in parentheses are t-statistics.

The standard SAR model specification (LeSage and Pace 2009) is as follows:

$$y = \rho W y + \alpha \iota_n + X\beta + \varepsilon \qquad (12)$$

where y is an n by 1 vector of (continuous) response variables (such as home values), ρ is a scalar measuring the degree of spatial autocorrelation, W represents an n by n spatial-weight matrix, ι_n is an n by 1 vector of ones (α is the parameter for this constant terms),and X is an n by K matrix of covariate attributes. The error term, ε, is assumed to follow an iid normal distribution, $N(0, \sigma_\varepsilon{}^2 I_n)$. For a binary probit variation of this standard SAR model, a latent variable $y*$ is introduced in place of y, and actual discrete outcome y equals 1 if $y*>0$ and 0 otherwise (Ozturk and Irwin 2001, Lacombe et al. 2009, Hoshino 2009).

Parameter estimates can be conducted using maximum likelihood estimation (MLE) methods. But LeSage and Pace (2009) find that MLE is subject to computational difficulties; they cited findings from Beron and Vijverberg (2000), reporting that their SAR probit application experienced estimation times on the order of "hours" – for just a 49-observation problem. Thus, estimation was achieved here using the Bayesian procedure proposed by LeSage and Pace (2009). Table 9 presents these results, where y equals 0 for parcels remaining in undeveloped status, and 1 once developed (including residential, commercial, service/recreational, office and industrial uses).

Table 9. Results of SAR Binary Probit Model for Parcel Development

Variable	*Coefficient*	*StdDev*	*p-value*
Constant	-0.881	1.157	0.227
DistCBD	0.070	0.020	0.000
DistMnrArt	-0.119	0.033	0.000
DistMajArt	0.346	0.093	0.000
DistFwy	0.040	0.039	0.168
Slope	-0.019	0.008	0.011
Access	0.053	0.390	0.449
Perim-to-Area	7.602	1.276	0.000
Area	-0.010	0.005	0.040
ρ	0.273	0.028	0.002
AIC= 5,468.4			

Parameter estimates of the SAR probit model highly resemble those of the probit GWR in this binary set-up. However, distance to freeway and accessibility lost significance in the SAR specification while having remarkable t statistics in the GWR model. The estimate of the spatial parameter, ρ, is 0.273, indicating a relatively weak spatial autocorrelation. In addition, the GWR probit model yields a lower AIC than the SAR probit (when modeling the developed/undeveloped response), suggesting a better fit for the far more flexible GWR technique, as one might well expect. While the SAR approach greatly simplifies interpretation, much of the spatial relationship comes down to a single parameter, ρ. The GWR method allows for local regressions, thus accommodating a variety of potential spatial variations. However, GWR methods do not offer a single interpretation on the effect of variables and can make model application to new locations (where a local regression has not been performed) challenging if not impossible.

CONCLUSIONS

This paper applied GWR techniques to Travis County data in Texas, yielding a series of 33,912 and 29,812 MNL model runs for models of physical/shape change and land use change, respectively. The first model was used to analyze parcels‘ merging and subdivision activity between 2003 and 2008, while the latterwas used to anticipate land use change (from an undeveloped state) across five use categories over the same time period. The nearest 1,000 neighboring points were used in each case, with a weight matrix based on the tri-cube weighting specification. To counter the identification problems emerging from Hessian inversions due to heavily biased response (in land use outcomes), all developed land use types were later collapsed into one category, allowing for comparison of a binary SAR probit’s and a GWR probit’s outputs.

The results from the 5-level MNL GWR model of land use change indicate a spatial interesting pattern of various covariates’ influence on land use development. Proximity to the region’s CBD tends to have a positive impact on the development office space, with an average elasticity of 2.2, but reduce the likelihood of undeveloped parcels becoming residential, commercial or industrial in nature, everything else constant.

Distances to the three types of roadways were estimated to exert varying influences. For example, residential development exhibits a tendency to avoid proximity to freeways and major arterials in these data: a 1-percent increase in distances to these is estimated, on average, to increase the probability of residential development by 1.2% and 1.8%, respectively. In contrast, non-residential developments tend to cluster around these transportation facilities. Soil slope was found to be reduce the likelihood of non-residential development, as reflected in uniformly negative elasticities for the non-residential alternatives. However, in some areas, a steeper slope is considered an amenity by home developers (presumably for scenic reasons), as reflected by the positive and practically significant elasticities for about 4-percent of data points.

Coefficient estimates from the probit SAR model highly resemble those of the probit GWR model (for binary response). But *DistFwy* and *Access*, with t-statistics of 4.07 and -3.57 in the GWR model, do not seem to be significant in the SAR model. The probit GWR model’s average AIC was lower than that for the probit SAR model, which suggests that local regression tends to better account for spatial variation than spatial autoregressive processes (which heavily rely on a single spatial parameter, ρ). In conclusion, the binary-probit GWR model seems to surpass the binary-probit SAR model in anticipating development. A comparison between such methods (GWR and SAR) in a multinomial setting is of interest, and hopefully methods and code will eventually exist to estimate the SAR version for large-scale data sets like those used here. It also would be useful to exhaustively or strategically search for more optimal N values (to determine the nearest-neighbor rule), and to work with discrete-response spatial data sets of various types (e.g., soil types, vehicle-type choices, and vegetative cover types).

As noted, the model results can be applied in a variety of ways, for various estimates of interest. For example, (hypothetically) increasing all undeveloped parcel sizes by one-quarter acre is estimated to result in a 14 percent reduction of newly developed parcels (from the observed value of 15,004 transitions to 12,903), over the 5-year period. Simply a two-percent decrease in all parcels‘ distances to freeways, major arterials and then minor arterials, separately/in turn, is predicted to result in 16, 15, and 14 percent reductions (in all undeveloped parcels‘ developing), respectively.

In general, the results of this work suggest that GWR-facilitated local-area regressions can work reasonably well with spatially rich, discrete-response data sets, such as those found across regions at the parcel level. The GWR MNL model used here appears to capture a variety of behaviors. Such methods offer planners and modellers the potential for intersting longer-run, dynamic predictions, thereby facilitating transportation and land use planning and policy evaluations.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Foundation Award SES-0818066 for supporting the data collection and research work for this paper. Mrs. Annette Perrone provided valuable editorial support.

REFERENCES

Anselin, L., Florax, R.J.G.M. and Rey, S.J. (2004) Advances in Spatial Econometrics. Methodology, Tools and Applications. Springer-Verlag Berline Heidelberg.

Atkinson, P., German, S., Sear, D. and Clark, M. (2003) Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis* Vol. 35 No. 1: 58-82.

Beron, K.J. and W.P.M. Vijverberg (2000) Probit in a Spatial Context: A Monte Carlo Approach in Advances in Spatial Econometrics L Anselin, and R. Florax (eds.) Heidelberg: Springer-Verlag.

Chakir, R. and Parent, O. (2009) Determinants of Land Use Changes: a Spatial Multinomial Probit Approach. *Papaers in Regional Science* Vol. 88(2): 327-344.

Greene, W.H. (2003) *Econometric Analysis 5th Edition*. Prentic Hall. Upper Saddle River, N.J.

Fotheringham, S. (2003) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd., West Sussex, England.

Ghosh, D. and Manson, S. (2008) Robust Principal Component Analysis and Geographically Weighted Regression: Urbanization in the Twin Cities Metropolitan Area (TCMA). *URISA Journal* 20 (1): 15-25.

Hadayeghi, A., Shalaby, A., and Persaud, B. (2009) Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. *Accident Analysis & Prevention.* Vol. 42 Issue 2:676-688.

Haider, M. and Miller, E.J. (2000) Effects of Transportation Infrastructure and Location on Residential Real Estate Values. *Transportation Research Record 1722*: 1-8.

Huang, H. (2010) Chief Data Developer, Social Explorer. Email correspondence, on July 24.

Iacono, M., Levinson, D., and El-Geneidy, A. (2008) Models of Transportation and Land Use Change: A Guide to the Territory. *Journal of Planning Literature* 2008 22: 323-340.

Lacombe, D.J., Shaughnessy, T.M., and Holloway, G.J. (2009) The Dual Spatial Autoregressive Probit Model with an Application to the 2001 Congressional Farm Bill. Presented at the Regional Research Institute, West Virginia University,
Morgantown, West Virginia, March 12th, 2009.

LeSage, J. (1999) Applied Econometrics Using MATLAB. Accessed July 24th, 2010.

LeSage, J. and Pace K. (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC. Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300. Boca Raton, FL 33487-2742.

Lichstein, J. W., Simons, T.R., Shriner, S.A. and Franzreb, K.E. (2003) Spatial Autocorrelation and Autoregressive Models in Ecology. *Ecological Monographs* 72(3): 445-463.

Lin, H., Kang, S.L., Espey, M. and Allen, J. (2005) *Modeling Urban Sprawl and Land Use Change in a Coastal Area- A Nueral Network Approach*. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005.

Luo, J. and Nagaraj, K. (2008) Modeling Urban Growth with Geographically Weighted Multinomial Logistic Regression. *Proceedings of SPIE,* the International Society for Optical Engineering Vol. 7144, 71440M.

Manson, S. M. (2000) Agent-Based Dynamic Spatial Simulation of Land-use/Cover Change in the Yucatán Peninsula, Mexico. Presented at the 4th International Conference on Integrating *GIS and Environmental Modeling GIS/EM4*, Banff, Canada.

McFadden, D. (1978) *Modeling the Choice of Residential Location*. In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist et al. Amsterdam: North Holland Publishers. .

McMillen, D.P. and McDonald, J.F. (1999) Land Use before Zoning: The Case of 1920's Chicago. *Regional Science and Urban Economics* 29(4): 473-489.
McMillen, D.P., Professor of Economics, University of Illinois. Phone Conversation. July 20th, 2010.

Miller, H. (2004) Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers* 94(2): 284–289.

Millington, J., Romero-Calcerrada, R., Wainwright, J. and Perry, G. (2008) An Agent-Based Model of Mediterranean Agricultural Land-Use/Cover Change for Examining Wildfire Risk. *Journal of Artificial Societies and Social Simulation* 11(4) 4:32.

Munroe, D.K., Southworth, J. and Tucker, C. M. (2002) The Dynamics of Land-Cover Change in Western Honduras: Exploring Spatial and Temporal Complexity. *Agricultural Economics* 27(3): 355-369.

Nakaya, T., Fotheringham, S., Brunsdon, C. and Charlton, M. (2010) Geographically Weighted Poisson Regression for Disease Association Mapping. *Statistics in Medicine.* Vol. 24 Issue 17: 2695-2717.

Niemeier, D.A. (1997) Accessibility: An Evaluation Using Consumer Welfare. *Transportation* Vol. 24: 377-396.

Ognev-Himmelberger Y., Pearsall, H. and Rakshit, R. (2009) Concrete Evidence and Geographically Weighted Regression: a Regional Analysis of Wealth and the Land Cover in Massachusetts. *Applied Geography* 29(4): 478-487.

Ozturk, E. and Irwin, E.G. (2001) Explaining Household Location Choices Using a Spatial Probit Model. Explaining Household Location Choices Using a Spatial Probit Model. Paper presented at the 2001 Meeting of the Agricultural and Applied Economics Association. Accessed July 21st, 2010. http://ageconsearch.umn.edu/bitstream/20626/1/sp01oz01.pdf.

Páez, A., Uchida, T., and Miyamoto, K. (2002) A General Framework for Estimation and Inference of Geographically Weighted Regression Models: Location-Specific Kernel Bandwidths and a Test for Locational Heterogeneity. *Environment and Planning* A 34 (4): 733–754.

Páez, A. (2006) Exploring Contextual Variations in Land Use and Transport Analysis Using a Probit Model with Geographical Weights. *Journal of Transport Geography* 14: 167–176.

Parker, C. D., Berger, T. and Manson, M.S. (2001) Agent-Based Models of Land-Use and Land-Cover Change. Report and Review of an International Workshop October 4-7, Irvine, California, USA. Accessed June 28th, 2010. http://www.globallandproject.org/Documents/LUCC_No_6.pdf

Parker, D. C., and Filatova, T. (2008) A Conceptual Design for a Bilateral Agent-Based Land Market with Heterogeneous Economic Agents. *Computers, Environment and Urban Systems* Vol.32, No. 6, 454-463.

Platt, R. (2004) Global and Local Analysis of Fragmentation in a Mountain Region of Colorado. *Agriculture, Econsystem and Environment* 101: 207-218.

Sermons, W., and Seredich, N. (2001) Assessing Traveler Responsiveness to Land and Location Based Accessibility and Mobility Solutions. *Transportation Research Part D6:* 417-428.

Srour, I., Kockelman, K. and Dunn, T. (2002) Accessibility Indices: A Connection to Residential Land Prices and Location Choice. *Transportation Research Record No.* 1805: 25-34.

Hoshino, T. (2009) GMM Estimation of Spatial Autoregressive Probit Models: An Analysis of the Implementation of the District Planning System in Japan. Paper presented at the 3rd World Conference of the Spatial Econometrics Association, in Barcelona, Spain. Accessed July 21st, 2010. http://www.ub.edu/sea2009.com/Papers/28.pdf

Tobler W. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46(2): 234-240.

Verburg, P.H., Van Eck JRR, Nijs TCM and Dijst MJ. (2004). Determinants of Land-Use Change Patterns in the Netherlands. *Environmental and Planning B* 31: 125-150.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke M. and Ulfarsson, F. (2003) Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics* Vol. 3.1: 43-67.

Wang, X. (2007) Capturing Patterns of Spatial and Temporal Autocorrelation in Land Use/Land Cover Change Data: An Austin Case Study. PhD dissertation, Department of Civil, Architectural and Environmental Engineering. University of Texas at Austin.

Wang, X., and Kockelman, K. (2009a) Bayesian Inference for Ordered Response Data with a Dynamic Spatial-Ordered Probit Model *Journal of Regional Science* 49 (5): 877-913.

Wang, X. and Kockelman, K. (2009b) Application of the Dynamic Spatial Ordered Probit Model: Patterns of Land Development Change in Austin, Texas. *Regional Science* 88 (2): 345-366.

Wang, X. and Kockelman, K. (2009c) Application of the Dynamic Spatial Ordered Probit Model: Patterns of Ozone Concentration in Austin, Texas. *Transportation Research Record* No. 2132: 13-24.

Yu, D. (2007) Modeling Owner-Occupied Single-Family House Values in the City of Milwaukee: a Geographically Weighted Regression Approach. *GIS Science & Remote Sensing* 44(3):267-282.

Zhao, F. and Park, N. (2004) Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record* 1879: 99-107.

Zhou, B. and Kockelman, K. (2008) Neighborhood Impacts on Land Use Change: A Multinomial Logit Model of Spatial Relationships. *Annals of Regional Science* 42 (2): 321-340.

Zhou, B. and Kockelman, K. (2010) Land Use Change through Microsimulation of Market Dynamics: An Agent-based Model of Land Development and Locator Bidding in Austin, Texas. Proceedings of the 56th Annual Meetings of the N. American Regional Science Assoc. International, in San Francisco; and under review for publication in *Transportation Research Record*.