

1  
2  
3  
4 **Lessons Learned in Developing and Applying Land Use Model**  
5 **Systems: A Parcel-based Example**  
6

7  
8 By

9  
10 Bin (Brenda) Zhou  
11 Graduate Student Researcher  
12 Department of Civil, Architectural and Environmental Engineering  
13 The University of Texas at Austin  
14 brendazhou@mail.utexas.edu  
15

16 Kara M. Kockelman  
17 (Corresponding author)  
18 Associate Professor and William J. Murray Jr. Fellow  
19 Department of Civil, Architectural and Environmental Engineering  
20 The University of Texas at Austin  
21 6.9 E. Cockrell Jr. Hall  
22 Austin, TX 78712-1076  
23 kcockelm@mail.utexas.edu  
24 Phone: 512-471-0210  
25 FAX: 512-475-8744  
26

27 To be presented at the 88th Annual Meeting of the Transportation Research Board, January 2009  
28 in Washington, DC, and under review for publication in *Transportation Research Record*  
29

30 November 2008

31 **ABSTRACT**

32 A variety of land use models now exist, driven by theoretical advances, data availability,  
33 enhanced computation, and new policy-making needs. This paper describes the process of  
34 developing and applying a disaggregate model system that seeks to simulate the subdivision and  
35 land use change of parcels, and the spatial allocation of households and employment across  
36 zones. Relying on multinomial logit specifications, random number generation, and a seemingly  
37 unrelated regression model with both spatial lag and spatial error components, the model  
38 development process was accompanied by a variety of challenges. Some issues are specific to  
39 the model presented here, but many are common for integrated transport and land use models. It  
40 is hoped that solutions to these issues and lessons learned offer useful insights for on-going  
41 improvements in land use modeling endeavors of all types.  
42

43 Comparisons of road pricing- and trend-scenario results for the Austin, Texas region  
44 highlight how policies may shape land and travel futures. While the road pricing policy did not  
45 alter land use intensity patterns in a significant way, it was forecasted to increase speeds across  
46 the region's network and reduce regional congestion levels.  
47  
48  
49  
50  
51  
52

## INTRODUCTION

Urban land use models (LUMs) seek to predict a region's future spatial distribution of households and employment, and provide key inputs to models of travel demand, emissions, and air quality. Integrated transport-land use models (ITLUMs) allow analysts to anticipate system response to new policies, preference functions, economic conditions and other scenarios. Though not nearly as complex as the human systems they seek to mimic, such model systems are very complicated. Such complication results in multiple challenges, and attendant abstractions result in many modeling limitations as well as prediction errors.

In this study, a new style of LUM is developed, to examine land use change at the parcel level while applying systems of equations for land use intensity (population and employment by type) across travel analysis zones (TAZs). This LUM was estimated for and then applied to the Austin-Round Rock Metropolitan Statistical Area (MSA) of Texas, under a business-as-usual scenario and a road pricing scenario (incorporating congestion tolls on freeways and carbon taxes on all roads).

As functionally distinct observational units, parcels lend themselves to disaggregate analysis with discrete responses for use type and subdivision. Models that emphasize parcel-level applications offer more behavioral realism (e.g., full-parcel development, rather than gridcell division) and enjoy significant potential in the land use modeling domain. Such models also allow meaningful investigation of the influence of highly local neighborhood conditions on parcel development. Past studies of neighborhood impacts (e.g. Verburg et al. 2004, Irwin and Bockstael 2004, Wang and Kockelman 2006) tend to use relatively coarse spatial resolution, focus on a single type of land use, or control for land cover categories identified from satellite images (rather than the actual land use). With the increasingly available parcel level data and the advancement of geographic information system (GIS) technology, parcel-level LUMs are not only advantageous but also feasible.

Parcel land development is often associated with changes of households and/or jobs, which are then aggregated to the level of TAZs in order to provide key inputs to travel demand models (TDMs). TAZs ensure reasonable representation of traffic networks and flows, but are rather arbitrary delineated. Anselin (1988) suggested that dependence is often present in cross-sectional data obtained using arbitrary spatial units (e.g., TAZ). Models without explicit treatment of these spatial dependencies may result in inappropriate inference. Recent spatial econometric advances (e.g. Das et al. 2003, Kelejian and Prucha 2004) allow modelers to disentangle such relationships, via spatial lag and spatial error processes.

This paper presents the logic of a largely parcel-based land use modeling system that accommodates various forms of spatial dependencies, in the land use change and intensity values. It also emphasizes the nature of key challenges encountered in model formation and application, their solutions, and lessons learned. It is hoped that these will shed light on future endeavors for developing more predictive yet practical LUMs.

## EXISTING LAND USE MODELS

With increasing computational power and theoretical advances, many operational LUMs have been developed. And several studies have summarized and compared existing models (e.g., Miller et al. 1998, PBQ&D 1999, US EPA 2000, and Dowling et al. 2005). The general consensus is that many limitations remain. Of course, to some extent there are "different horses

1  
2  
3 for different courses”, and the appropriateness and usefulness of any tool varies by context.  
4 Lemp et al. (2006) simply summarized four major theoretical constructs underlying the majority  
5 of LUMs: gravity allocation, cellular automata, spatial input-output, and discrete response  
6 simulation.  
7

8 In *gravity models*, regional transportation accessibility is core to the spatial allocation of  
9 jobs (by type) and households (by category). Zone-based specifications generally include lagged  
10 jobs and households, as well as some measure of land availability and land use conditions. Other  
11 influential factors, such as price adjustments, presence of built space, zoning restrictions, and  
12 topographic conditions are overlooked. Gravity models tend to use regional totals to adjust  
13 forecasts across all zones, and have been found to perform less well with disaggregate zone  
14 systems and/or sparse zone activity levels (PBQ&D 1999).

15 A representative gravity model is the Federal Highway Administration-sponsored  
16 Transportation Economic and Land Use Model (TELUM), which enjoys a user-friendly  
17 graphical user interface and is freely downloadable at <http://www.telus-national.org/index.htm>.  
18 However, its code is not shared, zone count is limited, and some key documentation is missing in  
19 its User Manual (2006) (e.g., parameter calibration, objective functions and land consumption  
20 variable definitions). A more flexible, open-source version of this model has been written in  
21 MATLAB, and is available at [http://www.ce.utexas.edu/prof/kockelman/G-](http://www.ce.utexas.edu/prof/kockelman/G-LUM_Website/homepage.htm)  
22 [LUM\\_Website/homepage.htm](http://www.ce.utexas.edu/prof/kockelman/G-LUM_Website/homepage.htm). This gravity model was applied to the Austin-Round Rock MSA  
23 (Zhou et al. 2008), and the forecasts only appear reasonable after imposing a series of rules  
24 (restricting excessive growth and declines in population and jobs at the zone level), suggesting  
25 that local knowledge and expert opinion may be needed to manually adjust gravity model  
26 forecasts.  
27

28 *Cellular automata* (CA) models are a class of artificial intelligence (AI) methods. Other  
29 AI methods include neural networks and generic algorithms, which also have been used to  
30 simulate and/or optimize land use change (Raju et al. 1998, Balling et al. 1999), but the CA-  
31 based SLEUTH model (Slope, Land use, Exclusion, Urban extent, Transportation and Hill  
32 shade) is the most widely applied (Clarke et al. 1997, Silva and Clarke 2002, Syphard et al.  
33 2005). It represents a dynamic system in which discrete cellular states are updated according to  
34 a cell’s own state, as well as that of its neighbors. However, SLEUTH relies on just five  
35 coefficients, and is calibrated in a rather ad hoc fashion<sup>1</sup>. While CA models may mimic many  
36 aspects of the dynamic and complex land use systems, they generally lack behavioral  
37 foundations to explain the process. Moreover, they emphasize land-cover type, not land use  
38 intensity, so post-processing is needed to generate employment and household count patterns  
39 (which are, of course, critical to travel demand modeling).

40 *Spatial input-output models* are used to anticipate the spatial and economic interactions of  
41 employment and household sectors across zones, using discrete choice models for mode and  
42 input-origin choices. Production and demand functions consider transport disutility between  
43 zones, and people (and generally freight) move from one location to another in order to  
44 equilibrate supply and demand. Representative models include TRANUS (see, e.g., Johnston  
45 and de la Barra 2000), PECAS (e.g. Hunt and Abraham 2003), and RUBMRIO (e.g., Kockelman  
46 et al. 2004). Trade-based spatial input-output models are most suitable for larger spatial units  
47

48  
49 <sup>1</sup> The model is calibrated by minimizing a variety of discrepancy measures, using historical data to initialize the runs  
50 and current data for comparison.  
51

1  
2  
3 (e.g., countries, regions, states and/or nations), so spatial resolution can be poor. Good trade and  
4 production data are also difficult to come by. It is worth noting that PECAS now includes a  
5 disaggregate sub-model for space development, to anticipate developer actions at the level of  
6 parcels or grid cells (see, e.g., PECAS 2007 and Hunt et al. 2008). This advance results in a  
7 hybrid of spatial input-output (for activity allocation) and microsimulation.  
8

9 Random utility maximization for *discrete choices* (McFadden 1978) is the basis of most  
10 microsimulation models. Waddell's UrbanSim (Waddell 2002, Waddell et al. 2003, Waddell and  
11 Ulfarsson 2004, Borning et al. 2007) simulates location choices of individual households and  
12 jobs, while anticipating new development on the basis of such models. In some contrast,  
13 Gregor's LUSDR (Land Use Scenario Developer) emphasizes fast model runs and the stochastic  
14 nature of results, seeking a balance between model completeness and practicality. Allocating  
15 groups of residential and business development on the basis of mostly multinomial logit (MNL)  
16 equations, LUSDR does not model price adjustments. The rationale behind utility maximization  
17 is defensible, but these choice-based models tend to require extensive data and consist of several  
18 submodels. The interactions among these model components make it hard to discern effects of a  
19 policy-decision variable, and uncertainties of one sub-model may be easily passed to other parts  
20 of the model system. In addition, numerous factors affect individual household and firm  
21 decisions, and these factors interact in complicated ways, often demanding some form of  
22 dynamic equilibration. For such reasons, opportunities for model improvement always exist. For  
23 example, UrbanSim does not (yet) tie households' workers to jobs or allow populations of jobs  
24 and workers to evolve. Many studies (Van Ommeren et al. 1999, Rouwendal and Meijer 2001,  
25 Clark et al. 2003, and Tillema et al. 2006) have suggested significant impacts of commute time  
26 (or cost) on residential and/or job site location decisions; and worker (and overall household)  
27 status can change rather rapidly.

28 While a variety of LUMs exist, the limitations of existing models are reasonably well  
29 understood. Many new modeling theories and approaches are emerging, including use of agent-  
30 based bidding (e.g. Zhou and Kockelman 2008c) and simulation of agent evolution and their  
31 dynamic behaviors (e.g., Moeckel et al. 2002). To make use of theoretical advances and recent  
32 available GIS data while illuminating challenges in LUM development, a very different approach  
33 was formulated here.

### 34 **ANOTHER APPROACH**

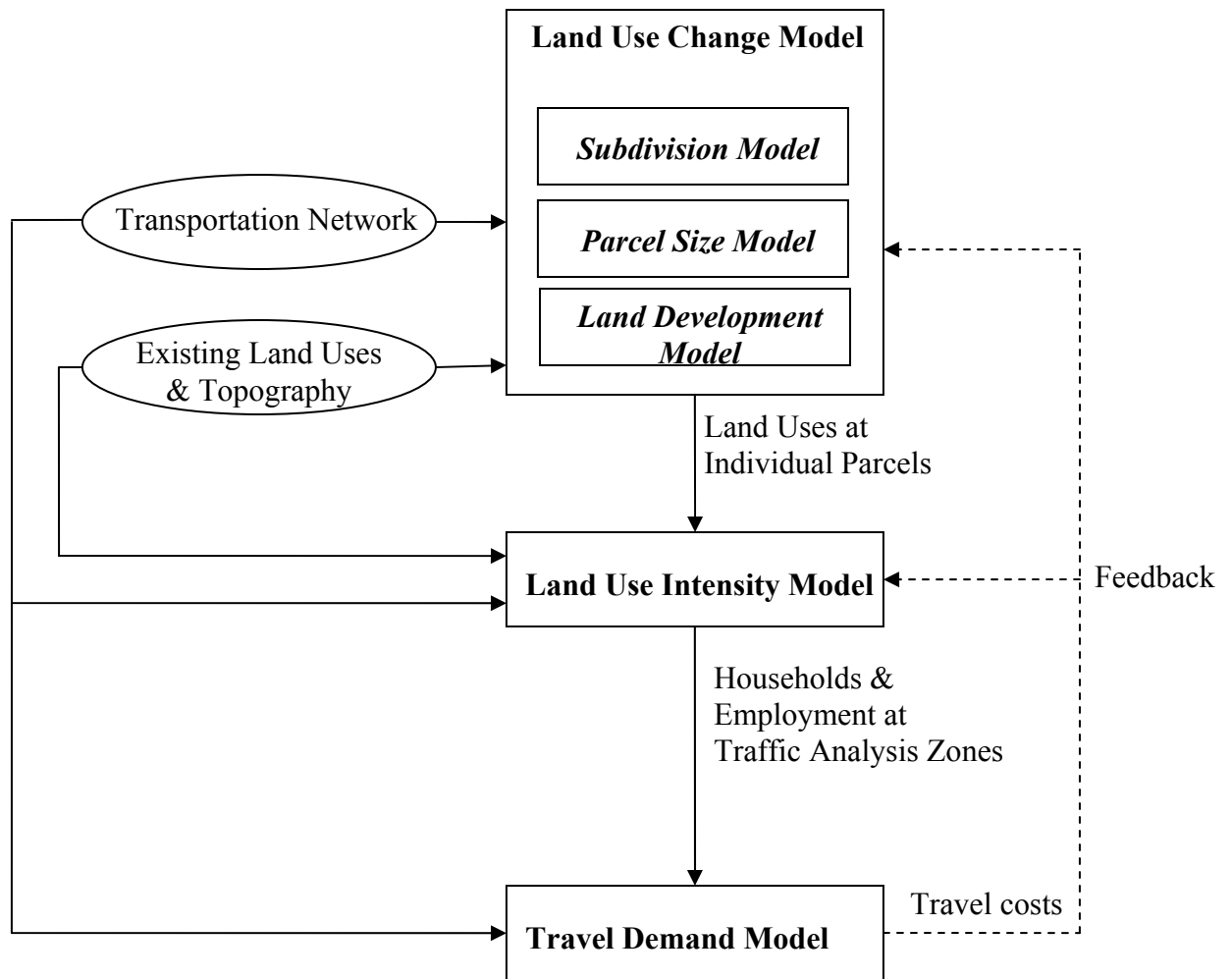
35  
36 First calibrated and then applied to the Austin-Round Rock MSA, this new LUM implements  
37 several recent advances in spatial econometrics, while exploiting the relatively recent availability  
38 of parcel-level data in Central Texas. Based on a model of *land use change* followed by a model  
39 of *land use intensity*, it is referred here as the LUC-LUI model.

40  
41 Figure 1 illustrates the relationships of various LUC-LUI model components. The *LUC*  
42 *model* anticipates how individual parcels evolve: whether an undeveloped parcel will divide into  
43 several smaller parcels during a pre-specified time interval (e.g., 5 years in this study), how big  
44 these subdivided parcels are likely to be, and what land use types will emerge on each (including  
45 undeveloped use). The likelihood of subdivision was simply modeled using a binomial logit and  
46 newly generated parcel sizes were determined using log-linear regression techniques. Land  
47 development on such previously undeveloped parcels was modeled using an MNL for various  
48 use alternatives (e.g., large lot single family, single family, multi-family, commercial or office,  
49 industrial, civic, and undeveloped). These alternatives are rather distinct; they do not include  
50  
51  
52

1  
2  
3  
4 mixed uses and neglect variations in building-floor densities. A variety of possible explanatory  
5 variables were considered, including attributes of individual parcels (e.g., size, shape, and slope),  
6 neighborhood conditions (nearby acres of different land uses [via a series of annuli, centered on  
7 each parcel in question]), and regional and local accessibilities terms (based on TDM results).  
8

9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

FIGURE 1 The Land Use Model System (LUC-LUI)



In the *LUC* model, only undeveloped parcels are considered (via a series of three sub-models). 56 percent of all undeveloped parcels (i.e., 22,400 out of 40,000 parcels) subdivided and/or changed use between 1995 and 2000 in the study area, while less than 1 percent of all developed parcels (just 1,100 parcels, out of 167,000 total) experienced such changes. In addition, as mentioned above, the *LUC* model focuses on land use types, rather than built space. Thus, focus on wholly new development seems reasonable.

Land use change at the parcel level is generally associated with an increase (or decrease) in activity, and the *LUI* model is designed to anticipate changes in the number of households and

1  
2  
3 employment (by type), using a seemingly unrelated regression (SUR) with two spatial processes,  
4 as follows:  
5

$$6 \quad \mathbf{y}_m = \rho_m \mathbf{W} \mathbf{y}_m + \boldsymbol{\beta}_m \mathbf{X}_m + \boldsymbol{\varepsilon}_m = (\rho_m, \boldsymbol{\beta}_m)' (\mathbf{W} \mathbf{y}_m, \mathbf{X}_m) + \boldsymbol{\varepsilon}_m = \boldsymbol{\delta}_m' \mathbf{Z}_m + \boldsymbol{\varepsilon}_m \quad (1)$$

$$7 \quad \boldsymbol{\varepsilon}_m = \lambda_m \mathbf{W} \boldsymbol{\varepsilon}_m + \boldsymbol{\xi}_m \quad (2)$$

8  
9 where  $m = 1:M$  and indexes the four equations for each zone (i.e., counts of households and  
10 basic, retail and commercial jobs),  $\mathbf{y}_m$  is an  $n$  by 1 vector of response variables for equation  $m$ ,  
11  $\mathbf{X}_m$  is a matrix of explanatory variables for equation  $m$  (including recent changes in land use  
12 areas, lagged counts of households and jobs, undeveloped land, land use balance and regional  
13 accessibility),  $\boldsymbol{\beta}_m$  is vector of parameters to be estimated,  $\rho_m$  is spatial lag autoregressive  
14 coefficient, and  $\lambda_m$  is spatial error autoregressive coefficient. The *LUC* and *LUI* model  
15 specifications are rather statistical in nature; due to space limitations, many of their details are  
16 described in Zhou and Kockelman (2008a, 2008b).  
17

18 The *LUC* and *LUI* models are sequentially applied, and outputs of the *LUC* model (i.e.,  
19 parcel-level land use changes) serve as key inputs to the *LUI* model (in the same model year).  
20 *LUI* effects filter to the *LUC* model in a downstream time step via regional and local  
21 accessibility variations, as anticipated by the TDM. Both land use models enjoy mechanisms to  
22 help ensure reasonable regional totals: The *LUC* model's alternative-specific constants were  
23 designed to reflect regional/five-county changes, and the *LUI* model's household and job  
24 forecasts were adjusted to match predetermined regional control totals (without any backwards  
25 adjustments to the *LUC* model). Issues with such adjustments are discussed later, in the  
26 Additional Lessons section.

## 27 DATA ISSUES

28 Data sets for detailed models of land use systems generally come from multiple sources, and the  
29 Austin *LUC-LUI* application is no exception. Here, the data sets include land use parcel maps,  
30 base-year counts of households (by category) and jobs (by type) at the TAZ level, transportation  
31 network details, and topographic data. Unfortunately, land use parcel data for the entire (five-  
32 county) MSA is limited (to a single year), requiring creative treatment of estimated parameters  
33 for reasonable results when the *LUC* model is applied to the entire region.  
34

35 In order to track the dynamics of parcel evolution, the *LUC* model requires parcel maps  
36 at two (or more) points in time. Only one (for the year 2005) could be obtained from the Capital  
37 Area Council of Governments (CAPCOG), who assembled the map using property appraisal  
38 data. Thus, the *LUC* model was estimated initially only for the City of Austin and its two-mile  
39 extraterritorial jurisdiction (ETJ) using 1995 and 2000 parcel maps. This model was then  
40 applied to the entire region's set of year-2005 undeveloped parcels. To ensure that region-wide  
41 forecast values matched target values<sup>2</sup> within a predefined tolerance (1 square mile per land use  
42 type in this study of a 4,280-square mile region), alternative specific constants were iteratively  
43 adjusted.  
44

45 As noted earlier, the *LUI* model estimates changes of household and job counts (by type)  
46 across five-year time steps, based on land use changes as well as lagged values of land use  
47 intensity, undeveloped land, land use balance and regional accessibility. It should be noted that  
48

---

49 <sup>2</sup> The region-wide "targets" were forecasted based on 2005 land use parcel map, as well as regional household and  
50 job counts/targets for 2005 and 2010.  
51

1  
2  
3 the LUI model, which relies on a SUR model with two spatial processes and a spatial weight  
4 matrix, cannot be easily extended to a larger region, with new zones added, nor some subset of  
5 zones<sup>3</sup>. For this reason, year 2000 land use conditions had to be backcasted for each TAZ, to  
6 ensure that the application set of zones matched the model estimation. Such is a limitation of  
7 various spatial econometric techniques, meriting one’s attention well before final model  
8 specification. Of course, this same guidance is true in models of all types: one should understand  
9 all modeling objectives (including policy tests and outputs of interest), data set limitations and  
10 any methodological constraints before finalizing the specification. Especially, data deficiencies  
11 present many other issues for model formulation and application, as described below.

### 12 **ADDITIONAL LESSONS**

13  
14 As alluded to above, further details on the LUC-LUI model’s specification, parameter estimates,  
15 and application results can be found in Zhou and Kockelman’s (2008d) report. For purposes of  
16 this paper, a more interesting question is how does such a model perform in practice? What  
17 deficiencies and challenges emerge, and can these be resolved?

#### 18 **Limits on Development and Model Extrapolations**

19  
20 Left unchecked, models calibrated on the basis of past trends in land use change can lead to  
21 overdevelopment or underdevelopment in the future. This is particularly true in a case where  
22 more centrally located data points (e.g., parcels within the City of Austin and its ETJ) determine  
23 the specification and their parameters are “transferred” to the regional setting. While this  
24 particular scenario was avoided here in the first time step – thanks to early parameter adjustment  
25 based on a five-year target – such adjustments may be of little use over the longer term, and  
26 model estimates can take on a life of their own.

27  
28 If targets are not embedded naturally into the model, post-processing is generally used,  
29 requiring heroic and unsatisfying assumptions (such as proportional adjustment of all zones’  
30 values) in order to hit regional control totals. Here, the LUC and LUI models are sequentially  
31 applied, and forecasts of the LUI model had to be adjusted in order to ensure reasonable regional  
32 totals. Parcel subdivision and land use change could be undertaken (in random order across  
33 undeveloped zones, as is presently the case) until a certain target of total land in each use is met.  
34 But this does not guarantee that application of the subsequent LUI model will then result in  
35 reasonable household and job counts for each model year. Totals by type can get out of synch,  
36 and modelers are left with little or no recourse.

37  
38 Meeting control totals is a tricky issue that deserves great care. While models like  
39 UrbanSim and LUSDR allocate jobs and households up to pre-specified targets, in each  
40 category, thus avoiding the control-totals issue. They do not enjoy endogenous determination of  
41 such quantities and rely on random order in agent assignments in each model year, thus  
42 neglecting within-year bidding competition for scarce space. In order to match regional control  
43 total expectations for Austin, the LUI model results were post-processed, in two steps. First,  
44 households and jobs were added or removed uniformly across the entire MSA (in proportion to  
45 TAZ sizes) in order to maintain their relative values. Next, unreasonable forecasts were removed  
46 using simple rules: if a zone was forecasted to have a negative count, zero values were assigned.

47  
48 <sup>3</sup> The issue in prediction essentially is that any zone’s (best) prediction depends, in a simultaneous way, on those of  
49 its neighbors, as with standard time-series data. However, prediction for a single equation has been proposed in  
50 more standard spatial auto-regressive systems (Pace and LeSage 2008).

1  
2  
3 Adjustment of zone-level LUC model results (by scaling land acreages up or down, according to  
4 use type) could have been pursued as well. Whichever post-processing technique is used,  
5 however, the options are unsatisfying.  
6

7 Related to the idea of parameter transferability is the idea of model extrapolation – over  
8 time and space. As models move forward in time, past data sets and their parameter estimates  
9 are less and less likely to apply. Preferences evolve, along with incomes, household sizes,  
10 technologies, energy prices, and various other key inputs. As regions expand, variables and  
11 estimates that made good sense for the original region lose their value. One example is sub-  
12 center nucleation, as a previously one-center region begins to enjoy multiple centers. Another  
13 example is the variable of distances to the nearest highway. Since paired land use data sets were  
14 only available for the City of Austin and its ETJ, this variable’s maximum was originally 1.2  
15 miles. For application of the resulting model to the larger region, the distance grew to 40.3 miles  
16 in certain cases, potentially resulting in values that dominated model results. Here, this variable  
17 was capped at 1.2 miles across all locations; no great options generally exist in such cases.  
18 Proper data sets need to be present from the start in order to avoid these kinds of questionable  
19 assumptions relatively late in the modeling process.  
20

### 21 **Implications of Model Specification for Purposes of Policy Analysis**

22 While decision-making support is the primary purpose of most LUMs, it often ends up being  
23 difficult to code a variety of *land use policies* in a particular model. Consider the example at  
24 hand: certain parcels can be removed from LUC consideration (due to an urban growth  
25 boundary, for example), their land use alternatives may be restricted (due to zoning, for  
26 example), or the attractiveness of their alternatives adjusted (due to subsidies for different land  
27 use types, for example); however, a sophisticated model of counts, using a spatial system of  
28 equations that share information in their error terms is not easily adapted to cases where certain  
29 zones are not allowed to experience growth in certain land use intensity values (due to zoning or  
30 other constraints) or some subset of values are otherwise pre-determined (due to prior knowledge  
31 of near-term development, for example). Spatial econometric tools are still emerging, and this  
32 challenge may one day be resolved, but in the meantime it represents a hurdle here that was  
33 unfortunately unforeseen.

34 The effects of *transportation policies* will largely depend on the accompanying TDM’s  
35 specification, and the LUM linkage. For example, variable-road pricing policies will require that  
36 time and cost metrics appropriately impact trip time of day, mode, and destination choice  
37 decisions, and that network assignment results feed back into trip distribution and other travel  
38 decisions (for consistency in sub-model assumptions and system-wide equilibration). While  
39 travel demand model specifications are largely outside the scope of this paper, it is very  
40 important to consider specification of the transportation-land use linkage, to ensure that  
41 transportation policies can affect land use patterns, if we believe that land use patterns respond in  
42 some way to transportation system conditions. Typical practice is to have regional accessibility  
43 terms feature in zone attractiveness equations for new development. In less simplistic and more  
44 spatially detailed LUMs (e.g., 150 m grid cells commonly used in UrbanSim or the parcels used  
45 here), more local transportation system attributes can be used. Here, they included parcel-based  
46 number of transit stops within 0.5 mile, network travel times to the region’s CBD, and distance  
47 to the nearest freeway.  
48  
49  
50  
51  
52



1  
2  
3  
4 Other policies of interest also exist. The proposed LUM does not consider land price  
5 signals, and is insensitive to fiscal policies, such as differential property taxation and subsidies.  
6 This could be remedied, to some extent, by controlling for neighboring parcels' land prices in the  
7 LUC Model (to anticipate land value impacts on development decisions) and directly modeling  
8 property valuation (for price predictions across parcels in each time step).

### 9 **Inclusion and Significance of Explanatory Variables**

10 This issue simply refers to the fact that various explanatory factors are not controlled for in the  
11 specification (e.g., soil quality, property tax rates, the presence of views, and school district  
12 scores). Ideally, more meaningful factors impacting land development decisions should be  
13 included, to enhance model flexibility in application. Of course, this brings us back to the very  
14 fundamental issue of data availability: someone will first need to assemble such variables for all  
15 zones/locations, trusting that these vary a fair bit across zones (which is unlikely with some  
16 variables, like construction costs, in many cases), and then hope that they emerge as statistically  
17 significant and with intuitive signs in the estimated parameter set. There are simply no  
18 guarantees that the data acquisition efforts will pay off. And it generally is enough work to  
19 acquire more basic information (like parcel location, current and past land uses, network and  
20 demographic variables, neighborhood land use conditions for each parcel, and so forth);  
21 expending constrained resources to acquire variables that may or may not offer much to the  
22 model is a real risk.

23  
24 In addition, when integrating a TDM and LUM, no transportation-system variables are  
25 guaranteed to have a practically or statistically significant effect on location choice or land  
26 development patterns. Location choice is a complex decision. Land markets reflect the  
27 interacting decisions of multiple actors, and developers, households and firms are not entirely  
28 rational or well informed of their options. Moreover, reliance on cross-sectional data sets for  
29 location choices (very common in practice, due to data availability) misses true move decisions,  
30 and can lead to anomalous results (e.g., negative coefficients on regional accessibility variables),  
31 reflecting the fact that many jobs and households are entrenched in past location choices and  
32 built spaces, rather than being able to choose anew, based on rational decision-making processes,  
33 where access is thoughtfully considered. In practice, simple Euclidean distance-to-CBD metrics  
34 are often far better predictors of land values, land use change, and land use intensity than more  
35 meaningful and rigorous measures of urban form (e.g., Kockelman 1997, Bina and Kockelman  
36 2006).

37 Such facts of modeling are unfortunate, but real. Many policies of interest may simply  
38 not be testable because the data sets exhibit no reasonable response to these variables or the  
39 variables do not vary over observational units to begin with (which is regularly the case with  
40 macroeconomic indicators [like interest rates] or potential policy variables that affect an entire  
41 region's agents all at once [such as gas prices]). Moreover, many desirable data sets simply do  
42 not exist. For example, when formulating the parcel-based LUM discussed here, zoning variable  
43 simply does not exist for the entire region.

### 44 **MODEL APPLICATION**

45  
46 The above discussion illuminates many modeling issues present in various LUM paradigms,  
47 including the one used here. In applying the LUC-LUI model, coupled with a relatively standard  
48 TDM, one can illuminate even more potential issues and modeling challenges. Here, the  
49  
50  
51

integrated LUM-TDM system was applied to anticipate the year-2030 distribution of households and jobs and travel conditions across the Austin-Round Rock MSA.

To better appreciate model sensitivities and the potential implications of different policies, two scenarios were examined: a business-as-usual (BAU) or trend scenario and a road pricing scenario. The *BAU scenario* assumes that development trends observed over the five-year calibration will continue, and no new policies are imposed. The *road pricing scenario* combines congestion pricing with a gas (or carbon) tax. A congestion charge was set to equal the implicit cost of marginal delay (imposed per added vehicle-mile-traveled, assuming a \$6.75/person-hour value of travel time) on all freeway segments in the network, and the carbon tax was assumed to be 4.55 cents per mile<sup>4</sup> on all links in the network. These two scenarios were simulated through year 2030 at five-year intervals.

### Simulating Behavioral Randomness

Since land use changes are discrete in nature, it is very natural to model such behavior via Monte Carlo simulation of MNL alternatives. To achieve something similar in the LUI model, a Cholesky decomposition (see, e.g., Greene 2001) of the SUR model system’s covariance matrix (shown as Table 1) was used to generate appropriate multivariate normal random numbers. To ensure greater comparability between scenarios, the same series of random numbers were used in each scenario (though different seeds were used in different model years).

TABLE 1 Covariance Matrix of Residuals in the Land Use Intensity (LUI) Model

	Households	Basic Jobs	Retail Jobs	Service Jobs
Households	2,2461	2,214	-251	824
Basic Jobs	2,214	170,054	346	-9,115
Retail Jobs	-251	346	23,968	7,013
Service Jobs	824	-9,115	7,013	135,772

Note: The model’s response variables are the *changes* in each of the four counts (households and three job types) in a five-year interval, and the number of observations (TAZs) is 1,245.

Year 2030 forecasts were generated based on the cumulative simulation results of five model years (2010, 2015, 2020, 2025, and 2030), which can average out some elements of randomness from time step to time step. Ideally, analysts simulate the results of such models over 50 to 100 times, to indicate the range and average of behavioral tendencies over time (as described, e.g., in Law and Kelton 2006 and Gregor 2007). In this case, however, the burden of computing neighborhood variables (at several annuli) for the LUC model and the time required to run TDM precluded multiple runs. (Each 25-year sequence of LUM and TDM runs required roughly 40 hours, not including manipulation and interpretation of results.<sup>5</sup>) This highlights the

<sup>4</sup> This value was based on assumptions that every gallon of gasoline sold at the pump is responsible for the emission of 26 pounds of carbon dioxide (U.S. EPA 2007), the cost of removing carbon from the atmosphere (or simply avoiding its production) is \$70/ton, and average fuel economy is 20 miles per gallon of gasoline.

<sup>5</sup> Kakaraparthi and Kockelman’s (2008) recent experiences with UrbanSim are better, as that expertly coded package requires just about about 15 hours for LUM predictions across the three-county study region, at 150 m gridcell resolution, over the 25-year period. However, each interaction with a TDM requires another 3 hours (for data transfer and model runs in TransCAD). Thus, for the same 25-year period, the entire process entails roughly 30 hours (before final output manipulations are performed).

1  
2  
3 benefits of simpler models; while they generally come at a cost of spatial resolution as well as  
4 behavioral and/or econometric sophistication, they do offer proficient programmers an  
5 opportunity to batch run multiple scenarios (simulating variation in any number of assumptions,  
6 including policy variables, parameter estimates, and input assumptions) and offer stakeholders  
7 far more information on potential model outcomes.  
8

9 Only through model application did a clear issue emerge in predictions:  
10 heteroskedasticity in job and household counts (across zones) was unaccounted for in the LUI  
11 model specification. Largely as a result of this, a couple very small TAZs (with areas of just  
12 0.013 square mile) ended up with what appear to be excessive counts (achieving year 2030  
13 household densities of 87 and 47 households per acre). Related to this are issues apparent in the  
14 original data sets, provided by the regional MPO: four TAZs ostensibly lost more than 500  
15 households between 2000 and 2005 (with the biggest loss at 1645), seven lost more than 2000  
16 basic jobs (8062 was the biggest loss) while five gained more than 2000 basic jobs (the biggest  
17 gain was 9118), and six TAZs lost more than 2000 service jobs (maximum of 4413) while four  
18 gained more than 2000 such jobs (maximum of 6277). Gaining thousands of households or jobs  
19 over a five-year period is questionable, and losing thousands is unrealistic during a period when  
20 the region is growing. While the econometrically sophisticated approach outperformed other  
21 existing model specifications (Zhou and Kockelman 2008b), such data inconsistencies can take a  
22 toll on model performance (overall  $R^2 = 0.320$ ). Roughly, 68% of overall count variation could  
23 not be explained by the model, contributing to extreme values in simulation forecasts.  
24

### 25 **Determining New Parcel Shapes**

26 Another valuable lesson emerges in application of the parcel subdivision model. Application of  
27 the log-linear regression model of new-parcel sizes is straightforward, but providing parcel  
28 shapes is another matter.

29 This study used ArcGIS and MATLAB software to rasterize the parcel maps into 240 feet  
30 grid cells and then converted these maps into ASCII files for use (as a matrix) in MATLAB.  
31 MATLAB assembled 240 ft cells from left to right in each subdividing parcel, and then up to  
32 down, in order to hit predicted new-parcel sizes. This arbitrary approach resulted in new parcels  
33 with a strong east-west orientation, rather than more natural patterns of parcel formation. New  
34 parcel shaping is a difficult issue to resolve using basic mathematical techniques, but certain  
35 parallels for better models may exist in engineering studies of breakage, for example.  
36

### 37 **SCENARIO COMPARISONS**

38 Figure 2a and 2b show the forecasted land use intensity patterns across TAZs in year 2030 for  
39 the business-as-usual and road pricing scenarios. Both sets of predictions suggest that  
40 households and jobs tend to cluster in urban areas and along regional freeways. The road pricing  
41 scenario resulted in very similar land use patterns, indicating that the increased travel costs  
42 (averaging just 6 cents/mile) may not alter location choices, though this policy does have a  
43 significant impact on travel behavior predictions.  
44  
45  
46  
47  
48  
49  
50  
51  
52

FIGURE 2a Household and Employment Densities in Year 2030 (Business-As-Usual Scenario)

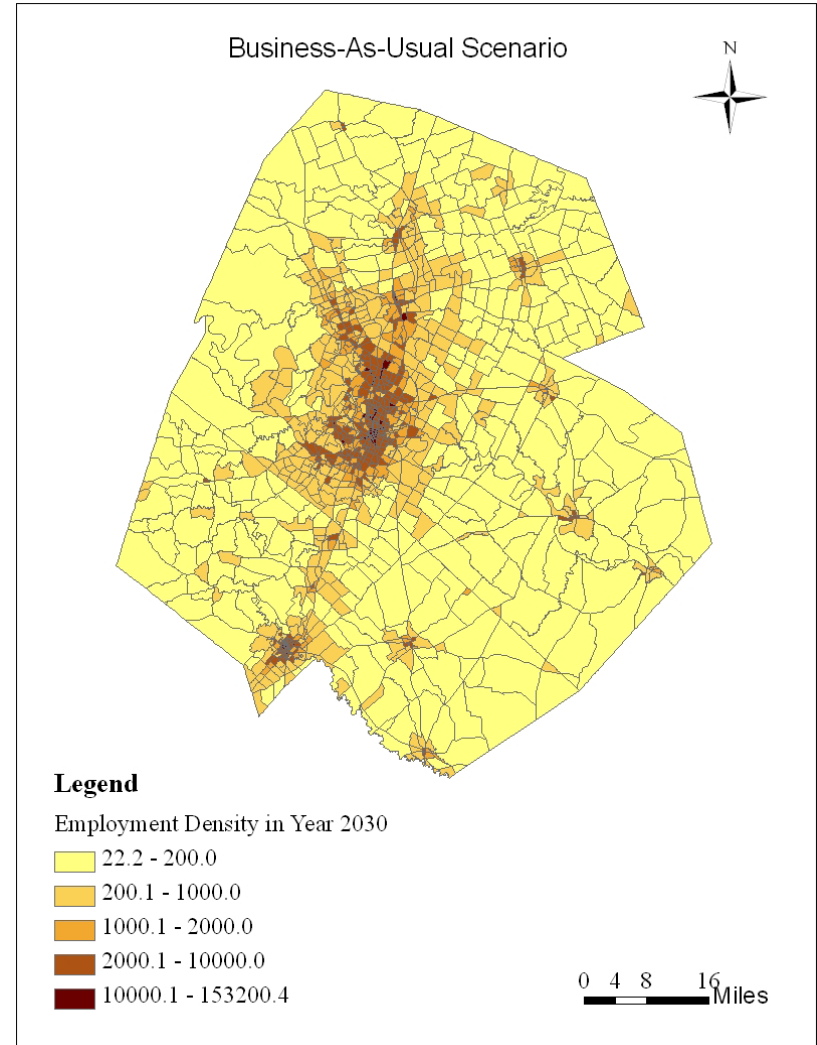
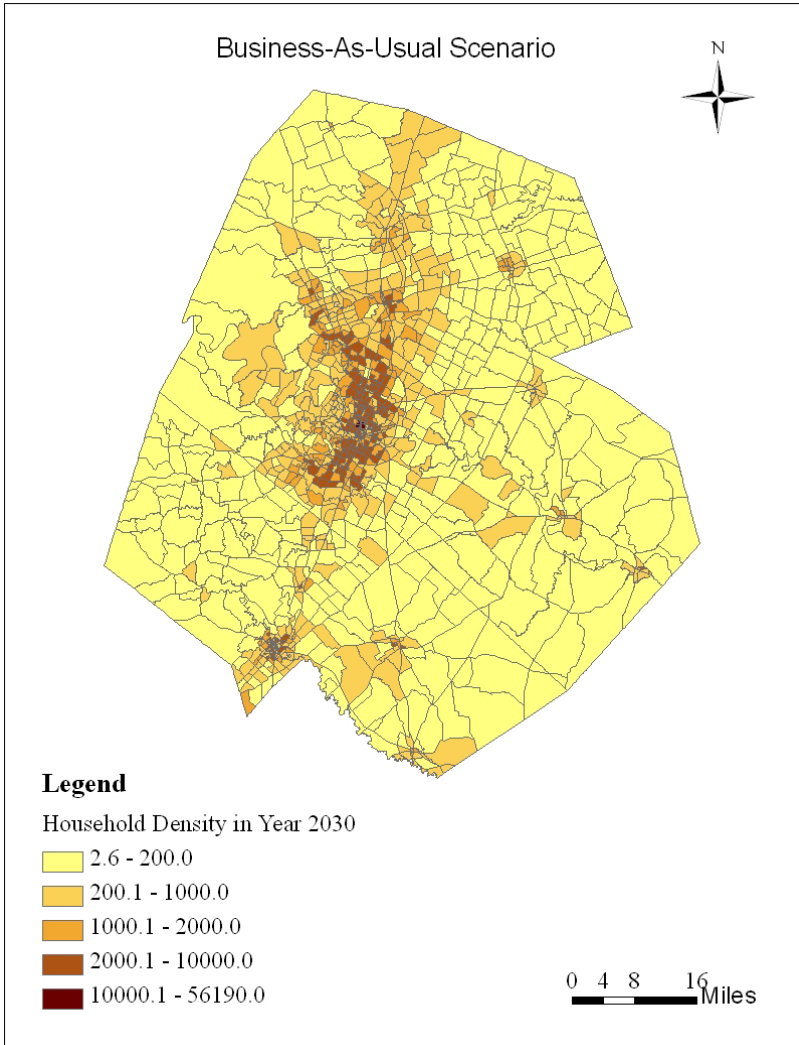
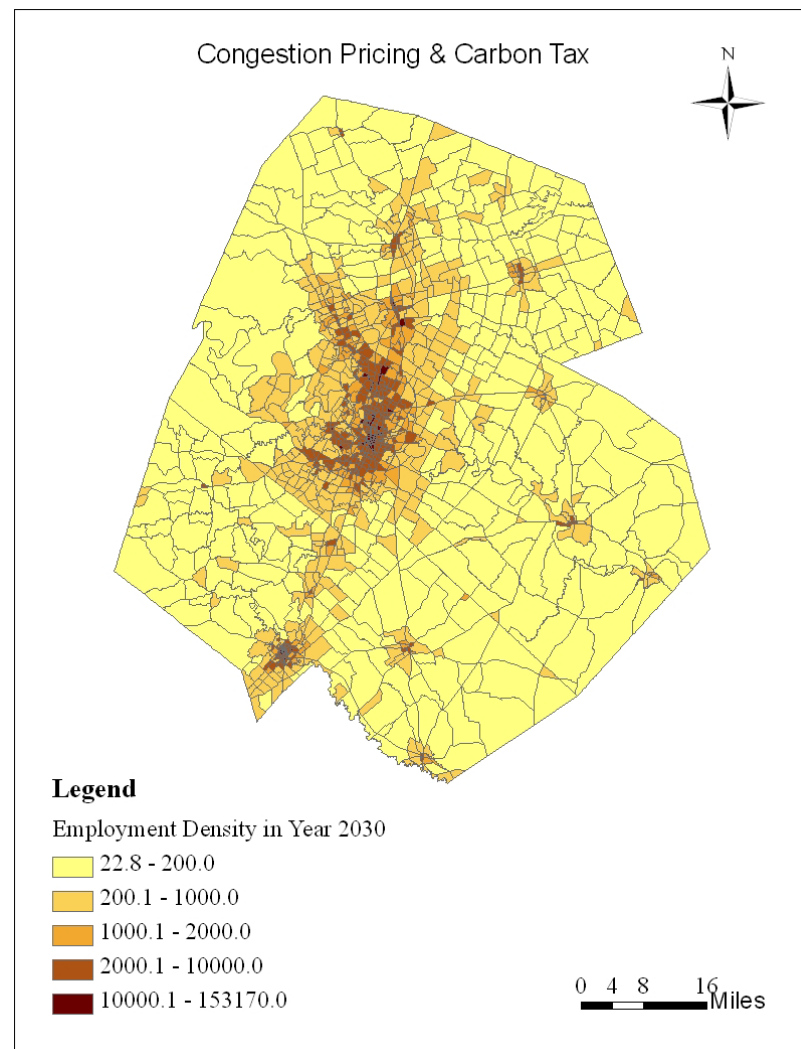
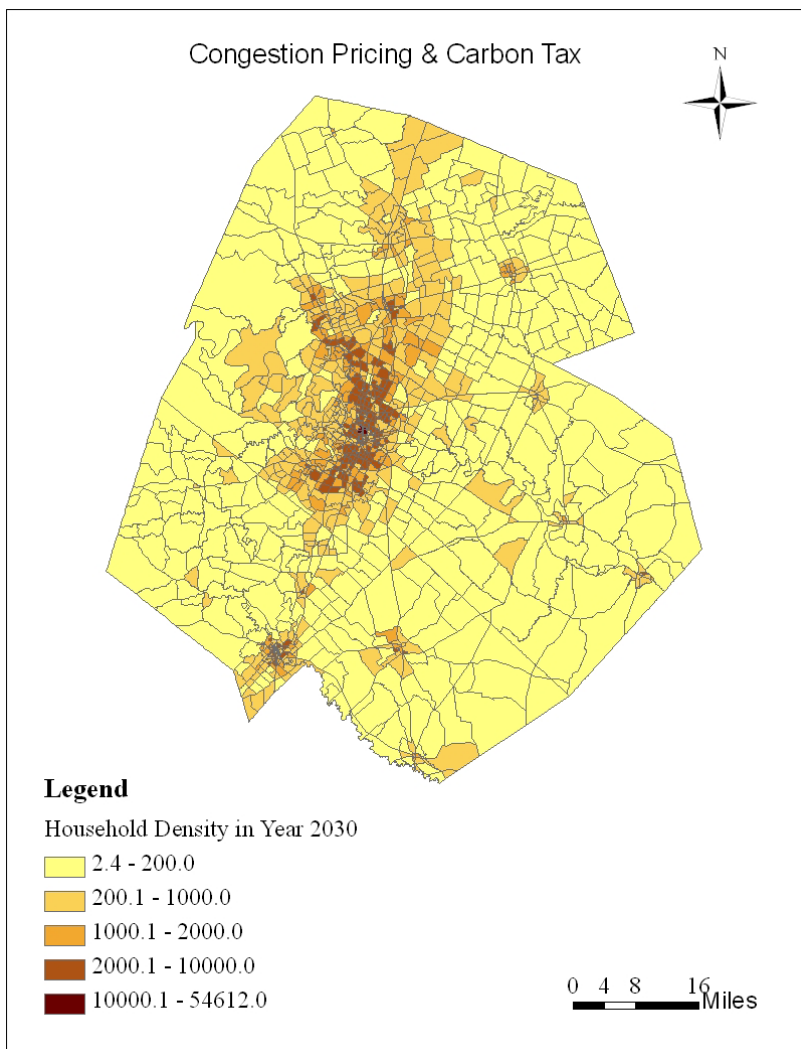


FIGURE 2b Household and Employment Densities in Year 2030 (Congestion Pricing & Carbon Tax Scenario)



1  
2  
3  
4 The count-weighted average densities for the BAU and road pricing scenarios were found  
5 to be similar: 1,568 and 1,532 households per square mile, and 6,391 and 6,375 jobs per square  
6 mile, respectively. Such minimal feedback from TDM to LUM was also found in recent  
7 application of the same scenario in a gravity-based LUM (integrated with the same TDM) for the  
8 same region (Zhou et al. 2008), supporting the notion that moderate increases in travel costs may  
9 not be reflected in land markets. The sensitivity of LUMs to TDM certainly deserves further  
10 investigation, using various modeling approaches and policy tests.

11 Of course, if a LUM is relatively insensitive to TDM inputs, modelers can skip the  
12 trouble of model integration and simply run the LUM forward to the future year of interest  
13 (holding accessibilities and most other transportation system variables constant), running the  
14 TDM only at the final stage. Such decisions will depend on the size and scope of transportation  
15 system changes under the scenario – and, in the hopes of urban system modelers everywhere,  
16 actual behavioral tendencies.

17 Regional vehicle miles traveled (VMT) is forecasted to almost double over the 25-year  
18 period, reaching 83.7 million vehicle-miles per day under the BAU scenario; but just 71.0  
19 million/day under the road pricing scenario (a 15% reduction from the BAU total, or a 66%  
20 increase relative to base year VMT). VMT-weighted speeds and VMT-weighted volume-to-  
21 capacity ratios suggest that road pricing increases speeds about 6 percent across the region’s  
22 network during peak hours (from 50.7 to 53.9 miles/hour) while reducing average peak-period  
23 volume-to-capacity ratios by 18 percent (from 0.622 in the BAU scenario to 0.511 in the road  
24 pricing scenario).

## 25 **CONCLUSIONS**

26  
27 In order to illustrate the challenges and complexities inherent in LUM development and  
28 application, this study devised a new model of household and employment patterns, using logit  
29 models for undeveloped parcel subdivision and land use change, continuous models of new-  
30 parcel size, and a system of spatial equations for zone-level counts. While this new LUM  
31 utilizes recent advances in parcel data generation and manipulation, as well as recent innovations  
32 in spatial econometrics theory, like nearly all LUMs, it leaves much to be desired.

33  
34 While various deficiencies and challenges emerged during model formulation, calibration  
35 and application, most were largely resolved via careful model specification, investigation of  
36 early forecasts, and subsequent parameter adjustments. However, others remain (such as  
37 heteroskedasticity in zone-level counts changes), resulting in certain facets of unrealistic  
38 prediction. As in nearly all LUMs, the data sets used here come from several sources, and data  
39 deficiencies are responsible for many of the issues encountered. Nevertheless, ensuring that  
40 models have a way to hit reasonable control targets (for total population and employment, for  
41 example) can be key to avoiding later (generally highly unsatisfactory) post-processing of such  
42 critical outputs. Similarly, ensuring natural mechanisms for matching supply and demand (of  
43 land and built space) can avoid a host of later problems.

44 Policy analysis is generally core to the development of LUMs, and policy objectives  
45 should have a significant impact on model specification, as well as estimation and application  
46 methods. Not all needs can be anticipated from the start, however, and faster model run times  
47 enjoy the advantage of facilitating diagnostic tests for model errors and timely correction of  
48 deficiencies. In general, there is much to be learned from the actual process of developing new  
49 model specifications, but all LUMs are likely to remain imperfect for the foreseeable future.

1  
2  
3 Nevertheless, there is much of interest and much to gain from such journeys, offering at least  
4 some insight for respective futures.  
5

## 6 **ACKNOWLEDGEMENTS**

7 The authors thank Dr. James LeSage at Texas State University for advice on prediction  
8 techniques when spatial dependent variables include both sample and ex-sample observations,  
9 Ms. Annette Perrone for her administrative assistance, and several anonymous reviewers for their  
10 comments. We also want to thank the U.S. Environmental Protection Agency STAR Grant  
11 project for financially supporting this study under Project 831183901, "Regional Development,  
12 Population Trend, and Technology Change Impacts on Future Air Pollution Emissions."  
13

## 14 **REFERENCES**

- 15 Balling, R.J., J.T. Taber, M.R. Brown, and K. Day. (1999) Multi-objective Urban Planning Using  
16 Genetic Algorithm. *Journal of Urban Planning and Development*, Vol. 125, No. 2, pp.  
17 86-99.
- 18 Borning, A., P. Waddell, and R. Förster. (2007) UrbanSim: Using Simulation to Inform Public  
19 Deliberation and Decision-Making. *Digital Government: Advanced Research and Case*  
20 *Studies*, R. Traunmueller, et al. (eds.), Springer-Verlag, New York.
- 21 Clarke, K.C., S. Hoppen, and L. Gaydos. (1997) A self-modifying cellular automaton model of  
22 historical urbanization in the San Francisco Bay area. *Environment and Planning B:*  
23 *Planning and Design*, Vol. 24, pp. 247-261.
- 24 Clark, W.A.V., Huang, Y., and Withers, S. (2003) Does commuting distance matter? commuting  
25 tolerance and residential change. *Regional Science and Urban Economics*, Vol. 33, No. 2,  
26 pp. 199-221.
- 27  
28 Das, D., H.H. Kelejian, and I.R. Prucha. (2003). Finite sample properties of estimators of spatial  
29 autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82,  
30 1-26.
- 31  
32 Dowling, R., R. Ireson, A. Skabardonis, D. Gillen, and P. Stopher. (2005) *National Cooperative*  
33 *Highway Research Program (NCHRP) Report 535: Predicting Air Quality Effects of*  
34 *Traffic-Flow Improvements: Final Report and User's Guide*. Transportation Research  
35 Board, Washington, D.C.
- 36  
37 Greene, W. *Econometric Analysis*. Upper Saddle River: Prentice-Hall, 2000.
- 38  
39 Gregor, B. (2007) The Land Use Scenario Developer (LUSDR): A Practical Land Use Model  
40 Using a Stochastic Microsimulation Framework. *Transportation Research Record*, 07-  
41 0438, pp. 93-102.
- 42  
43 Hunt J.D., and J.E. Abraham. (2003) Design and application of the PECAS land use modelling  
44 system. The 8th Computers in Urban Planning and Urban Management Conference,  
45 Sendai, Japan. <http://www.ucalgary.ca/~jabraham/Papers/pecas/summary.html>.
- 46  
47 Hunt J.D., J.E. Abraham, D. De Silva, M. Zhong, J. Bridges, and J. Mysko. (2008) Developing  
48 and Applying a Parcel-Level Simulation of Developer Actions in Baltimore. Proceeding  
49 of the 87th Annual Meeting of the Transportation Research Board, Washington D.C.,  
50 January 2008.  
51  
52

- 1  
2  
3  
4 Irwin E, and N. Bockstael. (2004) Land use externalities, open space preservation, and urban  
5 sprawl. *Regional Science and Urban Economics*, Vol. 34, pp. 705-725.
- 6 Johnston, R.A., and T. de la Barra. (2000) Comprehensive Regional Modeling for Long-Range  
7 Planning: Linking Integrated Urban Models and Geographic Information Systems.  
8 *Transportation Research Part A: Policy and Practice*, Vol. 34, No. 2, pp. 125-136.
- 9 Kakaraparthi, S., and K. Kockelman. (2008) An Application of UrbanSim to the Austin, Texas  
10 Region: Integrated-Model Forecasts for the Year 2030. Paper under review for  
11 publication in *Transportation Research Record*.
- 12 Kelejian, H.H., and I.R., Prucha. (2004). Estimation of simultaneous systems of spatially  
13 interrelated cross sectional equations. *Journal of Economics* 118, 27–50.
- 14 Kockelman, K. (1997) The Effects of Location Elements on Home Purchase Prices and Rents:  
15 Evidence from the San Francisco Bay Area, *Transportation Research Record* No. 1606,  
16 pp. 40-50.
- 17 Kockelman, K. M., L. Jin, Y. Zhao, and N. Ruiz-Juri (2004) Tracking Land Use, Transport, and  
18 Industrial Production using Random-Utility-Based Multizonal Input-Output Models:  
19 Applications for Texas Trade. *Journal of Transport Geography*, Vol. 13, No. 3, pp. 275-  
20 286.
- 21 Law, A.M., and W.D. Kelton. (2006) Simulation Modeling and Analysis, 4th Edition, McGraw-  
22 Hill Higher Education.
- 23 Lemp, J., B. Zhou, K.M. Kockelman, and B. Parmenter. (2006) Visioning vs. Modeling:  
24 Analyzing the Land Use-Transportation Futures of Urban Regions. Forthcoming in  
25 *Journal of Urban Planning and Development*.
- 26 McFadden, D. (1978) Modeling the Choice of Residential Location, in A. Karlquist et al. (ed.),  
27 *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam.
- 28 Miller, E. J., D.S. Kriger, and J.D. Hunt. (1998) Integrated Urban Models for Simulation of  
29 Transit and Land-Use Policies: Final Report. Transit Cooperative Research Project.  
30 National Academy of Sciences.
- 31 Moeckel, R., C. Schürmann, and M. Wegener. (2002) Microsimulation of Urban Land Use. The  
32 42nd European Congress of the Regional Science Association, Dortmund.  
33 <http://www.raumplanung.uni-dortmund.de/rwp/ersa2002/cd-rom/papers/261.pdf>
- 34 Parsons Brinckerhoff Quade and Douglas (PBQ&D). (1999) *National Cooperative Highway*  
35 *Research Program (NCHRP) Report 423A: Land-Use Impacts of Transportation: A*  
36 *Guidebook*. Transportation Research Board, Washington, D.C.
- 37 Pace, R.K., and J.P. LeSage. (2008) Spatial Econometric Models, Prediction, *Encyclopedia of*  
38 *Geographical Information Science*, Shashi Shekhar and Hui Xiong (eds.), Springer-  
39 Verlag.
- 40 PECAS (2007) Theoretical Formulation: System Documentation Technical Memorandum 1.  
41 October 2007. (Received from John Abraham in March 2008)
- 42 Raju, K.A., P.K. Sikdar, and S.L. Dhingra. (1998) Micro-simulation of residential location  
43 choice and its variation. *Computers, Environment and Urban Systems*, Vol. 22, No. 3, pp.  
44  
45  
46  
47  
48  
49  
50  
51  
52



- 1  
2  
3 203-218.  
4  
5 Rouwendal, J., and Meijer, E. (2001) Preferences for housing, jobs, and commuting: a mixed  
6 logit analysis. *Journal of Regional Science*, Vol. 41, No. 3, pp. 475-505.  
7  
8 Silva, E.A., and K.C. Clarke. (2002) Calibration of the SLEUTH urban growth model for Lisbon  
9 and Porto, Portugal. *Computers, Environment and Urban System*, Vol. 26, pp. 525-552.  
10  
11 Syphard, A.D., K.C. Clarke, and J. Franklin. (2005) Using a cellular automaton model to forecast  
12 the effects of urban growth on habitat pattern in southern California. *Ecological  
13 Complexity*, Vol. 2, pp. 185-203.  
14  
15 Tillema, T., Ettema, D., and Van Wee, B. (2006) Road pricing and (re)location decisions of  
16 households. Proceeding of the *85th Annual Meeting of the Transportation Research  
17 Board*, Washington D.C.  
18  
19 U.S. Environmental Protection Agency (U.S. EPA). (2000) *Projecting Land-Use Change: A  
20 Summary of Models for Assessing the Effects of Community Growth and Change on  
21 Land-Use Patterns*. Washington, D.C., Report EPA 600-R-00-098.  
22  
23 U.S. Environmental Protection Agency (U.S. EPA). (2007). Lifecycle Impacts on Fossil Energy  
24 and Greenhouse Gases. Chapter 6 of *Regulatory Impact Analysis: Renewable Fuel  
25 Standard Program*. Washington, D.C., Report EPA 420-R-07-004,  
26  
27 User Manual: TELUM (Transportation Economic and Land Use Model) Version 5.0.  
28 <http://www.telus-national.org/telum/TELUMUserManual.pdf>. Accessed March, 2006.  
29  
30 Van Ommeren, J., Rietveld, P., and Nijkamp, P. (1999) Job moving, residential moving, and  
31 commuting: a search perspective. *Journal of Urban Economics*, Vol. 46, No. 2, pp. 230-  
32 253.  
33  
34 Verburg P.H., J.R.R. Van Eck, T.C.M. Nijs, and M.J. Dijst. (2004) Determinants of land-use  
35 change patterns in the Netherlands. *Environmental and Planning B*, Vol. 31, pp. 125-150.  
36  
37 Waddell, P. (2002) UrbanSim: Modeling Urban Development for Land Use, Transportation and  
38 Environmental Planning. *Journal of the American Planning Association*, Vol. 68, No. 3,  
39 pp. 297-314.  
40  
41 Waddell, P., A., Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. (2003)  
42 Microsimulation of Urban Development and Location Choices: Design and  
43 Implementation of UrbanSim. *Networks and Spatial Economics*, Vol. 3, No. 1, pp. 43 –  
44 67.  
45  
46 Waddell, P., and G.F. Ulfarsson. (2004) Introduction to Urban Simulation: Design and  
47 Development of Operational Models. In *Handbook in Transport, Volume 5: Transport  
48 Geography and Spatial Systems*, Stopher, Button, Kingsley, Hensher eds. Pergamon  
49 Press.  
50  
51 Wang, X. and K.M. Kockelman. (2006) Tracking Land Cover Change in a Mixed Logit Model:  
52 Recognizing Temporal and Spatial Effects. *Transportation Research Record* No. 1977,  
pp. 112-120.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

Zhou, B., and K.M. Kockelman. (2008a) Neighborhood Impacts on Land Use Change: A Multinomial Logit Model of Spatial Relationships. *Annals of Regional Science*, Vol. 42, No. 2, pp. 321-340.

Zhou, B., and K.M. Kockelman. (2008b) Predicting the Distribution of Employment and Households: A Seemingly Unrelated Regression Model with Spatial Dependence. Forthcoming in *Journal of Transport Geography*.

Zhou, B., and K.M. Kockelman. (2008c) Microsimulation of Residential Land Development and Household Location Choices: Bidding for Land in Austin, Texas. Forthcoming in *Transportation Research Record*.

Zhou, B., and K.M. Kockelman (2008d) *Predicting Land Development and Spatial Distribution of Households and Employment: An Application of a Novel Land Use Model System and Lessons Learned*. Internal Report for STAR Grant program project, sponsored by the U.S. Environmental Protection Agency, at The University of Texas at Austin. Available at <http://www.ce.utexas.edu/prof/kockelman/>.

Zhou, B., K.M. Kockelman, and J. Lemp. (2008) Transportation and Land Use Policy Analysis Using Integrated Transport and Gravity-based Land Use Models. Proceeding of the 88th Annual Meeting of the Transportation Research Board, Washington D.C., January 2009.

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

**LIST OF TABLES AND FIGURES**

TABLE 1 Covariance Matrix of Residuals in the Land Use Intensity (LUI) Model

FIGURE 1 The Land Use Model System (LUC-LUI)

FIGURE 2a Household Density in Year 2030 (Business-As-Usual Scenario)

FIGURE 2b Employment Density in Year 2030 (Business-As-Usual Scenario)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52