1998 ACSE Conference
Transportation, Land Use, and Air Quality: Making the Connection
Portland, Oregon – May 1998

Exploiting Parcel-Level GIS for Land Use Modeling

Paul Waddell[1] Terry Moore[2] and Sharon Edwards[3]

Abstract

This paper describes the use of parcel-level data in the development of UrbanSim, the Oregon Prototype Metropolitan Land Use Model, and its application to the Eugene-Springfield Metropolitan Area[4]. The exploding use of Geographic Information Systems to map and manage tax parcel databases is a widely noted trend, but these massive databases have yet to be fully exploited in the development of new land use and transportation models. This paper explores a case study in the exploitation of a parcel-level GIS database in Lane County, and its integration with a business establishment inventory, to develop an internally consistent disaggregate database for use in calibrating the Oregon Prototype Metropolitan Land Use Model. The focus of the paper is on the challenges involved in using data of this level of spatial detail, and on the analysis that has been done to develop these data for use in the land use model.

Introduction

The policy motivations for developing the Oregon Prototype Metropolitan Land Use Model are rooted in the Growth Management strategies being pursued by state and local governments within Oregon. Numerous policy questions that cross the boundaries of land use and transportation have been raised, and have proven difficult to answer with existing tools. Prime examples of such questions

[1] Associate Professor, University of Washington, Graduate School of Public Affairs, Box 353055, Seattle, WA 98195.
[2] ECONorthwest, Eugene, OR.
[3] Lane Council of Governments, Eugene, OR
[4] This work was funded by the Oregon Department of Transportation as part of the Transportation and Land Use Model Integration Project.

include the impact of changes in the Urban Growth Boundary around a metropolitan area on land and housing prices and on the density and spatial pattern of development. Questions have also been raised about the effectiveness of alternative planning tools to direct more development into mixed-use development nodes, or to stimulate infill and redevelopment. Within a context of highly regulated land markets, with significant growth pressure and limited land supply, the use of highly disaggregate spatial information on land supply and its characteristics becomes a virtual requirement.

The availability of land for future development has recently become identified as a distinct research field, known as ‘land supply monitoring’ (see Godschalk, 1986). The concern about future availability of land for development is especially acute in areas that are actively managing growth through techniques such as Urban Growth Boundaries. What is not clear in much of this literature, however, is that land supply is a function not only of natural constraints and land use policies, but is an outcome of the interaction of market demand and developer behavior, subject to these constraints. For example, densification of development will occur as the dwindling quantity of vacant space faces increased competitive pressure for development. Similarly, we would expect competitive pressures to stimulate redevelopment of existing structures, making the measurement of available of land for development an even more elusive metric. These outcomes are a complex function of changes in prices and density of development, and require modeling techniques to sort out.

It is our view that questions concerning infill and redevelopment far exceed the scope of existing land use models, including DRAM/EMPAL (Putman, 1983), MEPLAN (Echenique, et al, 1990) and TRANUS (de la Barra, 1993). We would conclude that such issues not only require an explicit land market approach to modeling, but must also be based on parcel-level data and use of GIS.

Efforts to incorporate GIS into land use modeling to date have often been limited to adding a linkage to existing GIS tools, in order to quickly translate the tabular results of land use models into thematic maps (see Batty, 1992 and Langerdorf, 1995). While useful in quickly interpreting the results of a model forecast, this approach does not significantly extend our capability to address the questions outlined above.

The California Urban Futures Model (Landis, 1995) uses GIS processing to assemble layers of land use, jurisdictional boundaries, highways and other themes into a composite GIS layer that is used in a procedure to allocate residential development according to estimates of profitability. The representation of land is in this case a hypothetical ‘developable land unit’ that is not based on land ownership parcels, but still provides a high degree of spatial disaggregation. A later version of the CUF Model (Landis, 1997) uses a hectare grid as the spatial unit, and predicts the transition probabilities from one land use to another. In this research, Landis has shown the potential for using highly disaggregate GIS to analyze the impacts of alternative local land use policies on probable residential development patterns. As Landis has noted, however, the model does not include a land market, nor does it identify the actors in that market. By avoiding the use of parcel data, the database used in the CUF Model lacks information on quantities of built space and values of land and improvements that would be needed for a more market-based approach.

Perhaps the most intriguing research to date in the area of land use modeling is in the microsimulation approaches applied in the IRPUD model (Wegener and Spiekerman, 1996). Wegener and Spiekerman have described the potential for linking microsimulation techniques in urban modeling to highly disaggregate spatial data using GIS, though this description still does not exploit parcel-level data.

Over the past decade there have been revolutions in GIS, database, and computing technology, and massive investments have been committed to the development of parcel-level GIS databases within cities, counties, utilities and tax appraisal districts. One wonders why these data have not previously been exploited for their potential use in land use planning and forecasting. Several reasons may help explain this apparent oversight. First, the availability of parcel-level data with spatial coordinates is a very recent phenomenon, and many metropolitan areas still do not have complete coverage of the urban area. Second, access to such data has been hindered by a combination of proprietary data structures and often confused policies within public agencies about limiting access to enable cost-recovery for the significant expense of development and maintenance of these data. Third, the focus within the GIS community has heretofore been disproportionately placed on the database development and maintenance aspects of these parcel data, and not on their analytical potential. Fourth, the approaches to land use modeling have until recently been fairly aggregate in nature, and the theory to support more disaggregate approaches has not previously been embodied in an operational urban model.

In spite of these impediments, it appears that there are new opportunities to begin to more fully exploit the investments in parcel-level databases for use in land use modeling. The use of parcel-level data in the Oregon Prototype land use model provides a useful case study in both the promise and the limitations of such an application of these data.

Overview of the Model and Data Requirements

In this section we briefly describe UrbanSim, the Oregon Prototype Metropolitan Land Use Model, restricting the discussion to those aspects most relevant to understanding its use of parcel-level data. The model is based on a decision-maker approach to modeling urban development, and identifies households, businesses, developers, and governments as the major actors. Businesses and households make decisions about whether to move, and if they choose to move, which location to choose and what type of building to occupy. Developers make choices about which parcels of vacant land to develop, what type of building to construct, and at what density. Developers also select parcels with existing development as potential sites for redevelopment. These market-based choices interact, and are influenced by governmental decisions about infrastructure, and policies such as land use plans, Urban Growth Boundaries, and development impact fees.

The UrbanSim model uses discrete choice techniques to model the market choices of businesses, households, and developers, within particular policy scenarios that have been chosen by the user. The choices of location and building type are currently aggregated to the level of a Traffic Analysis Zone in order to maximize consistency with existing travel models. The developer component of the model, however, is based on the spatial unit of the land ownership parcel.

In order to address the policy questions outlined earlier, and in particular questions regarding infill and redevelopment, the developer model uses a database that integrates land ownership parcels with GIS layers representing environmental and policy constraints on development. This database integration process is similar to that done in the California Urban Futures Model, with the exception that this database includes land parcels. Once these data are integrated, then each vacant parcel is available for analysis in the developer model, with full information about its land value, size, location, land use plan designation, and whether it falls into any areas that would be subject to environmental or policy constraints. Constraints are imposed on land development by user-specified rules regarding the conversions of land allowed within each development plan category, the minimum and maximum density for new construction in each land use, and an adjustment to the maximum allowed density for parcels falling into each environmental constraint or within the Urban Growth Boundary.

For the analysis of redevelopment we need information on all parcels that have existing development, including the value of the land and any improvements on it. The key indicator used to identify parcels that might be targets for redevelopment is the ratio of improvement value to land value. Those parcels with the lowest ratios of improvement to land value are included in the developer model as competitive sites for development. Which parcels actually get built will depend on the relative profitability of the development, including the demolition costs and improvement costs for existing improvements, the ‘hard costs’ of new construction, development impact fees and other ‘soft costs’ of development, and the expected revenue from the development into alternative land uses.

There is another, less obvious, motivation for the use parcel-level GIS than its use in the developer component of the model. It results from the need to address the choices of building type made by households and businesses. What kind of housing a household will occupy is a distinct but related question from which zone they will choose to locate in. We need to be able to observe the choices made by households and businesses, not only of the zones of location choice, but simultaneously which building types they choose. A geocoding of households or businesses to the level of a Traffic Analysis Zone would not reveal the choice of building type within the zone. Since we assume that these are related choices and that businesses and households may substitute locations and building types on the basis of attributes such as cost and availability, we must develop input data that contains household and business geocoding to Traffic Analysis Zone and building type (land use). This requirement presents the most significant challenge for the database development.

Database Development

The database developed for the implementation of UrbanSim contains the following elements:

- Parcel GIS database containing approximately 73,000 parcels within the Lane COG planning area.
- Business establishment database for approximately 6,000 businesses within the planning area.
- A household database synthesized from census tabulations by census block group and the 5% Public Use Microdata Sample.
- GIS themes representing land development constraints, including the Urban Growth Boundary, high slopes, wetlands, floodways, stream buffers, and utility line easements.
- Zone to zone travel impedances from the travel demand model, for computing accessibility measures.
- Regional growth controls, to constrain the overall level of population and employment growth.

The parcel data used in this project originated from the Lane Council of Governments, and was available for 1994 in Arc/Info format. The database contained assessed land and improvement value, land use code, land use plan designation, and lot size. Although it included the number of residential housing units, the database did not originally contain the square footage of nonresidential buildings. Since the model explicitly accounts for the land market components of land, structures, and occupants, the square footage data was critical for the implementation of the model. Without square footage, the basic supply information for the nonresidential components of the model was missing. LCOG had conducted a study in Eugene in 1994 for developing input data for Trip Generation that included inventorying square footage on nonresidential parcels. A supplemental effort was undertaken by LCOG as part of this project, to collect comparable data for nonresidential parcels in Springfield and the balance of the planning area.

Other aspects of the parcel database that warrant further description include two aspects of its database design. First, ownership parcels were subdivided by LCOG into land use polygons wherever an ownership parcel contained more than one land use. The dilemma created by this design was that although the area and land use of these sub-parcel polygons were known, other available data such as land value, improvement value, and square footage of nonresidential space, were known or collected only at the level of the ownership parcel. This limited our ability to exploit the land use data, and for the purposes of the modeling, the ownership parcel was chosen as the basic unit of data. A second database design issue that became more prominent in subsequent data analysis, is that in instances in which a building or complex of buildings such as a mall crossed parcel boundaries, they were assigned entirely to one of the overlapping parcels. This would not have been a concern if the geocoding of businesses had been entirely consistent with the assignment of buildings to parcels. But as we shall discuss later, businesses were often assigned to the adjacent parcels from the one the building had been assigned to, creating the appearance of one parcel with a vacant building, and an adjacent parcel with a homeless business.

The business establishment database maintained by LCOG is based on the state employment commission employer database. It has known omissions regarding self-employment and proprietor establishments, and in most states has the additional problem that all of the employment within a multi-establishment business may be reported at the address of the headquarters or of an administrative office. LCOG has had an exceptional program in place since 1978, however, to obtain and geocode these business establishment data biennially, and to allocate headquarters employment to individual establishment locations. In addition, the geocoding procedures at LCOG are based on the use of a Master

Address File that links addresses to individual parcels. This meant that by address-matching the business establishment records, they were automatically linked to the land ownership parcels.

To develop a household database geocoded to the level of Traffic Analysis Zone and housing type, census data and parcel data were combined in an innovative synthesis approach. Procedures for imputing or synthesizing household data from combining geographically-detailed census tabulations with sample data such as the 5% Public Use Microdata have been described and used by various researchers in the context of developing microdata for models based on microsimulation (see for example Clarke, 1996). The techniques involved typically are based on either Iterative Proportional Fitting (IPF) or on reweighting of sample weights in a survey sample. The approach used in this project employed IPF to allocate household samples into census block groups using marginal tabulations by block group, and using parcel data on housing units by type by Traffic Analysis Zone and census block group to assign households to Traffic Analysis Zones.

The resulting database contained households and businesses stratified into groups, and geocoded to combinations of Traffic Analysis Zones and building types. Information on the parcel characteristics of each of these building objects was derived from the parcel database, including the quantity of housing units and nonresidential square footage, land and improvement values, and acreage. A link to the fully disaggregated parcel database is retained both in the initial database and throughout the execution of the UrbanSim model.

Database Diagnostics

Once the parcel database and business establishment data were linked, numerous diagnostics were conducted to assist in the analysis of the data. After linking these data, we were able to derive the following diagnostic calculations for each Traffic Analysis Zone and building type:

- Average square footage per employee
- Average density (units per acre or floor-area ratio)
- Improvement value per unit or nonresidential square foot
- Land value per acre

The preliminary analysis of these data suggested that there were some underlying problems in the data linkage. In particular, some zone and building type combinations had extremely low or high values on the average square footage per employee, suggesting some combination of missing square footage data, dubious employment levels for one or more businesses, or other possible explanations. We also found cases with odd density values, or improvement values per unit or square foot.

The difficulty in diagnosing these data issues became obvious as we attempted to sort out the competing explanations while looking at the linked and aggregated data. It was virtually impossible to determine the source of the problem without referring back to the original, disaggregate parcel and business establishment data. Ultimately, the cleanup of the database needed to be done at the level of individual businesses linked to individual parcels. At this level of analysis the problems, and therefore the solutions, generally became obvious. We attempted to prioritize the problems in order to maximize our use of limited time and resources, and recognizing that diminishing returns would quickly set in.

The following cases were the most common kinds of mismatches and problems that were encountered:

- Buildings or complexes of buildings that crossed parcel boundaries and were assigned to only one parcel, while one or more businesses in these buildings were assigned to the other parcel. This problem was addressed by grouping such parcels into 'superlots' to make the square footage and business geocoding consistent. This problem could have been further limited to addressing those cases that affected an assignment across building types or zones.

- Businesses that had more employment than could be reconciled with the existing square footage in the building. There were several legitimate reasons for some of these cases:

  - Businesses that had a large fraction of their workers working outside (but on-site), such as lumber or gravel companies.

  - Businesses that had most of their workers working off-site, such as construction contractors or temporary agencies.

  - Businesses that employ multiple shifts, such as medical facilities or some manufacturing plants.

  The solution to these cases depended on the particular circumstances, but the general rule that was applied is that the employment level needed to be consistent with daytime employment at the particular site. To achieve this, employment associated with off-site or evening-shift employment needed to be subtracted from the site.

After these steps were taken in cleaning the input data, calibration of the model proceeded much more effectively, and the more aggregate diagnostics done at the level of the zone and building type appear fairly clean. The total amount of effort expended in the diagnosis and cleanup of these data, while not insignificant, was certainly within the range to be expected for the development and application of any land use model. In addition, the cleanup was useful to LCOG for their other uses of these data.

Summary and Conclusions

Any land use or transportation model must rely on some measures of the spatial distribution of land, population, and employment. The underlying relationships between these three classes of data are often misunderstood, however, leading to serious errors in model estimation and application, even with a relatively high degree of aggregation. Employment data is almost exclusively available by industry grouping (SIC), but the use of employment data in trip generation is typically thought of in terms of land use categories such as warehouse, office or industrial space. Industries must ultimately be converted to land uses, and employees must be reconciled with square feet of space.

By integrating parcel-level data on buildings, land and improvement values, square footage, and businesses with their associated employment, a rigorous analysis of the consistency of these data was conducted. Using indicators such as the square footage per employee, the improvement value per square foot, and the ratio of building to land square footage, parcels with likely problems were flagged. A process involving GIS tools to integrate scanned aerial photography, parcel maps, and business locations was developed to correct inconsistencies and errors in these data, for use in subsequent model estimation.

The promise of parcel-level data for use in land use modeling and other rigorous analytical applications in planning and policy analysis appears to finally be within reach. The costs of parcel database development are being borne for totally unrelated reasons, and the timetable for development of such data are therefore driven by a calculus more oriented to land records management than to land use and transportation planning. But to the extent that it appears that investments in such parcel databases will continue to accelerate, and to the extent that questions continue to be raised regarding the efficacy of growth management strategies and the effective supply of land for future development, it would seem that the linkage of parcel-level GIS and land use modeling is a promising and perhaps inevitable evolution.

References

Clarke, G.P. 1996. "Microsimulation: an Introduction," in *Microsimulation for Urban and Regional Policy Analysis*. Pion: London.

De la Barra, Tomás. 1995. *Integrated Land Use and Transportation Modeling: Decision Chains and Hierarchies*. Cambridge University Press.

Echenique, Marcial, A.D.J. Flowerdew, J.D. Hunt, T.R. Mayo, I.J. Skidmore and D.C. Simmonds. 1990. *Transport Reviews* 10(4), (309-322).

Godshcalk, David R. 1986. *Land Supply Monitoring : a Guide for Improving Public and Private Urban Development Decisions.* Oelgeschlager, Gunn & Hain and the Lincoln Institute of Land Policy.

Putman, Stephen. 1983. *Integrated Urban Models: Policy Analysis of Transportation and Land Use.* Pion: London.

Wegener, Michael and Klaus Spiekerman. 1996. "The Potential of Microsimulation for Urban Models," in *Microsimulation for Urban and Regional Policy Analysis*. Pion: London.