Measuring Land Use Patterns for Transportation Research

Paul Mitchell Hess, Anne Vernez Moudon, and Miles G. Logsdon

Density and land use mix are focused on as the two primary variables for characterization of land use in transportation research. As commonly constructed, these variables do not capture well actual development patterns on the ground, thus obscuring a potentially strong relationship between land use and transportation behavior. To overcome these limitations, parcel-level data and geographic information system software were used to identify and measure attributes of land use. These data are at a level of resolution that closely corresponds to the spatial distribution of development patterns. A method for locating concentrations of medium-to-high-density housing and commercial development in suburban areas identified in previous research is described. The method includes the use of metrics derived from landscape ecology to model these development patterns and, specifically, their shapes and their functional and spatial mixes.

Land use has been recognized as one of the factors in transportation that can help reduce congestion and shift transportation choice away from single-occupant automobile travel. Researchers seeking to establish a causal relationship between land use, the built environment, and transportation behavior have identified a variety of issues that limit the results of research. These include the adequacy of existing data sets and the appropriateness of the variables used (1). This paper focuses on density and land use mix, the two primary land use variables shown to be significant in past research. The paper discusses the reasons why, as usually constructed, variables of density and land use mix poorly capture the types of land use patterns that theory suggests should be related to reduced automobile use. The paper first explains how commonly used geographic units of analysis are too large for use in the modeling of the spatial arrangement of land use and development patterns as they are found on the ground. It is argued that parcel-level data and geographic information system (GIS) software can be used for the type of fine-scale and spatially explicit data analysis necessary to capture land use patterns. Second, the paper models aspects of land use by drawing from a set of techniques developed by landscape ecologists to quantify land cover and landscape patterns. The land use and built form measures constructed provide a broader and richer set of techniques than have previously been available for quantification of the spatial structure of the built environment.

LIMITATIONS OF COMMONLY USED LAND USE VARIABLES

The results of research on relationships between urban form and travel behavior have been mixed, despite ambitious research and modeling efforts. Researchers have interpreted differently the weak results on how land use and the built environment may affect travel demand. In a recent review of the academic literature, Crane (2) concludes that not much can be said to policy makers as to whether the use of urban design and land use planning can help reduce automobile traffic. In contrast, Cervero and Kockelman argue that "research that enriches our understanding of how different elements of the built environment combine to shape travel behavior under different conditions is more imperative than ever" (1, p. 219).

The authors' previous research on development patterns suggests that weak relationships between land use and transportation stem in great part from the limited construction of variables that can be used to define density and land use mix. Specifically, commonly used land use variables suffer from the following shortcomings: a poor spatial boundary match between data and actual development patterns, the use of empirically untested proxy variables, and the use of measures that capture the heterogeneity rather than the complementarity of land uses.

Area and Boundaries of Spatial Units of Data and Analysis

One problem with establishing a strong relationship between land use and travel behavior is what geographers label the modifiable area boundary problem. This refers to the effects of spatial units of data and analysis on research results. For example, the use of different geographic units will yield different proportions of population living above certain densities. For King County, Washington (a part of the metropolitan Puget Sound region), for example, 1990 census tract-level data show that only 40,000 people are living at 25 people/ha or above. These data indicate that most of these higher-density tracts are located near the center of the city of Seattle. Yet, analysis of the same area with 1990 census data on the block level rather than the tract level shows that more than 400,000 people are living at 25 people/ha and above. Furthermore, these higher-density blocks are not only near Seattle, the largest and oldest central city in the county, but are spread throughout the urbanized part of the county. This example demonstrates that both the total number and the spatial distribution of people living at certain densities are highly sensitive to the size and boundaries of the spatial unit of data and analysis.

It would be less important to measure precisely the distribution of densities if the population was distributed evenly across the urbanized landscape. Yet, this is not the case, even in seemingly sprawling suburban environments. The authors' work has identified almost 100 concentrations of multifamily housing in the Puget Sound region; together these comprise 20 percent of the suburban population at densities that are more than twice those of the region (3). Furthermore,
these housing clusters are in close proximity to retail centers and schools, forming small, compact, and mixed-use areas that are less than 2.59 km² (1 mi²). Because they are smaller than typical units of analysis, the clusters have been overlooked in previous land use and transportation research. Census tracts and travel analysis zones (TAZs) typically average out these types of concentrations with adjacent lower-density development. Also, the boundaries of census tracts and TAZs normally act to divide land use concentrations that are located at the intersections of main roads, thus further diluting measures of land use intensity or mix.

Validities of Selected Variables

A second problem with establishment of strong relationships between land use and travel behavior may lie with the validities of research variables. Preoccupied with data availability and measuring such relationships, researchers have not dealt with how well variables are capturing the relevant attributes of the built environment. For example, block size (or intersection density) has been used as a proxy for pedestrian- and transit-oriented development patterns. Yet, this variable cannot be applied uniformly over the urban and suburban landscapes. Small blocks in pre-1930s urban environments correspond to areas with medium to high population densities, finely grained mixtures of uses, and highly connected pedestrian networks. In these areas, block size works as a reasonable proxy measure for pedestrian- and transit-oriented places. In contrast, however, block size in areas developed after World War II has a different relationship for places that may generate nonautomotive travel. Small blocks in these areas correspond to low-density, single-use subdivisions of detached housing. On the other hand, large blocks in areas developed after World War II correspond to areas with higher-intensity uses such as multifamily housing and retail development. Indeed, in many suburban places, the intensity of land uses and the likelihood that different uses occur in close proximity are inversely related to block size. Although these are not pedestrian-friendly places, post-war retail and multifamily environments are more likely and have actually been shown to generate more walking and transit trips than subdivisions with single-family homes. A regional basis, therefore, block size captures different intensities and mixes of land uses in urban and suburban areas. As a variable, block size will confuse research results because it does not have a linear relationship to environments that may generate walking and transit trips.

Land Use Mix and Measures of Heterogeneity

Commonly used measures of land use mix also have serious shortcomings. Two types of measures have been used in recent research efforts. Using tax assessors' land use data aggregated to the tract level, Frank and Pivo (6) used an entropy index as a measure of land use heterogeneity. The index, first used in a land use context by Cervero (7), compares the total area in different uses within a tract. It also compares the number of uses present in a tract with the total number of uses found in the data set. A tract with all uses present in the data set in the same areal proportion will have the maximum index value.

The researchers found some, albeit weak, statistical relationships between land use mix and travel mode. That stronger relationships were not found should not be surprising. First, as with measures of density, tracts are simply too large to capture the fine-scale variations in mix that may affect travel. Second, as shown in Figure 1, the measure does not capture the spatial interaction of different uses. Two tracts with the same overall proportion of uses will receive the same index value, even though one has uses in large zones and the other has uses in many smaller interspersed pockets. Also, the index weighs all categories of land use equally; thus, it does not take into account the fact that a tract split between office and industrial uses does not have the same travel implications as one split between residential and retail uses. Likewise, a tract with 30 percent of its area in residential uses and 70 percent in retail uses has different travel implications than a tract with the same uses in the reverse proportions, but this difference is also not captured.

Cervero and Kockelman developed a second measure of mix that they termed a dissimilarity index (1). The data consist of dominant land uses assigned to 1-ha squares of land. A value of dissimilarity is attached to each 1-ha square on the basis of the uses of the squares adjoining it. As the number of adjoining squares with uses that differ from that of the central square increases, the index value of the central square also increases. Note that the index is a measure only of whether adjoining squares are different (or not) from the central square and therefore is insensitive to the number of uses that are different from that of the central square.

Land use data at the hectare level capture a finer grain of land use mix than data aggregated to the tract level but are still unlikely to describe land use mix in urban areas developed before World War II. In these areas a hectare is roughly equivalent to a typical block size, and parcels in different uses are typically smaller. For example, hectare-level data will not catch land use mix on retail streets that run through a predominately residential area. Also, like the entropy index, the dissimilarity index measures only land use heterogeneity and the degree of overall mixing and does not distinguish between different types of mixed land uses and their travel implications.

![FIGURE 1 Three pairs of land use situations that have different transportation implications but that will be measured as the same by use of an entropy index. (Numbers indicate land use types.)](image-url)
Cervero and Kockelman reported that the dissimilarity index is a more powerful predictor of travel than the entropy index (1). It should also be noted that they attempted to address the limitation of the index by collecting a large set of additional primary data. However, those data did not help to strengthen the significance of the relationship between land use and transportation. Despite these limitations, the fact that both indices showed a statistically significant relationship between land use and travel serves to underscore the importance of developing more sophisticated measures of mix.

VARIABLES THAT CAN CAPTURE DEVELOPMENT PATTERNS

Land Use Complementarity

Land use complementarity is a more theoretically appropriate concept than the concept of heterogeneity or mix for transportation research. Complementary land uses are those that are likely to be linked by travel. Mixed yet noncomplementary land uses, such as agricultural and industrial land uses, will not be related to probable travel origins and destinations. Specific types of mixed uses will also affect individual trip frequencies, as well as the time of day and the days of the week when trips will likely occur. Trips between residential and office land uses, for example, will have different frequencies and temporal patterns than those between residential and retail land uses.

To capture these differences, three interrelated elements of land use must be expressed in measurements:

- Land use types, which serve as proxies for trip origins or destinations;
- Land use functional complementarity, which captures the presence of origins and destinations that are likely to be linked by travel; and
- Land use spatial complementarity, which ensures that functionally complementary land uses are within adequate proximity for the trip modes under consideration. Spatial complementarity in turn has two dimensions: one is the extent of the area where land uses are mixed and the other is the grain at which complementary land uses occur. These dimensions are discussed below.

Measures that incorporate these concepts will enable researchers to better test relationships between development patterns and travel.

Scale: Defining Area Extent and Grain

In measuring density and land use mix, researchers also need to explicitly consider issues of scale to develop appropriately powerful modeling techniques. Theoretical developments in other fields help clarify the issue of scale and its measurement. Cartographic geography, quantitative landscape ecology, and urban morphology have, among other fields, highlighted two basic measures of scale in an environment (8, 9). One is the extent of the area considered, and the other is the grain and level of resolution of the data used in the processes of both data collection and analysis. Measures of scale as extent and as grain or level of resolution can be applied to all objects or things because scale is pervasive through space and time. In analyzing urbanized regions, the researcher can select from many area extents: states, regions, local jurisdictions, centers, neighborhoods, and so forth. The corresponding units of spatial analysis can range from the individual parcels or subparcels of the land to the region. As illustrated below, advances in GIS technology now allow analyses to be carried out across large areas while using very small spatial units for data collection and analysis.

Making the distinction between the spatial unit of data collection and the spatial unit of analysis is important. As a general rule, the spatial unit of data collection should be as small as possible to capture variations in the phenomenon of interest. (The spatial unit of data collection is not related to data accuracy.) Along these lines, the paper has discussed how common measures of density and land use mix suffer substantial distortions because either the spatial unit of data collection (according to Cervero and Kockelman (1)) or the spatial unit of analysis (according to Frank and Pivo (6)) is larger than the patterns of land use on the ground.

MEASURING DEVELOPMENT PATTERNS WITH PARCEL-LEVEL DATA

To identify and analyze development patterns, the tax lot or parcel is used as the elementary spatial unit of data collection. In suburban areas in particular, parcels tend to be in single use and associated with specific types of buildings and site layouts, but parcels also capture many of the variations found in urban environments with their finely grained development patterns. The approach used here aggregates parcel-level uses into larger spatial units of analysis in spatially explicit ways that avoid many of the problems discussed above.

For the purpose of illustrating the approach, land use patterns that correspond to relatively small clusters of multifamily housing, retail uses, and schools are modeled. King County’s 600,000-parcel data set with 1998 tax assessor data coded to a GIS polygon coverage is used. The approach seeks to identify and analyze development patterns that may be related to certain types of travel, namely, home-based, nonwork pedestrian trips. Transportation researchers and planners analyzing the relationship between land use and household transportation behavior may modify the approach to model land uses around household locations. In general, research goals or policy questions will shape decisions relating to the coding and selection of land uses for analysis, the definition of geographic areas of analysis, and the application of particular metrics to describe land use patterns.

The approach draws on a large body of work developed in landscape ecology (8–11). Quantitative landscape ecologists have been concerned with quantifying the spatial arrangements of different land cover patches in landscapes. A landscape is defined as a bounded area of the Earth’s surface and corresponds to the extent of the area under analysis. A patch is conceived of as a contiguous area of relatively homogeneous land cover, for example, a spatially distinct area of mature forest or meadowland. Patches correspond to the spatial units of analysis (but not necessarily to the spatial unit of data collection). Finally, the structure of a landscape is quantified in terms of (a) the composition of the landscape, as defined by the presence and area of patches of different types or classes, and (b) the configuration of the landscape, as defined by the spatial distribution and arrangement of patches (9, 12).

Landscape ecologists measure land cover patterns to define the relationship between the dynamic structure of landscapes over time and biological and ecological processes. Many of their techniques are readily adaptable to the analysis of land use patterns, as opposed to land cover patterns. They can be used to test the relationships between these patterns and travel behavior.

The approach used here consists of three steps. First, parcels that are candidates for inclusion in clusters are selected to differentiate them from noncluster parcels. Second, the clusters are delineated
by grouping candidate parcels through an analysis of their geographic proximities. Finally, the internal compositions and configurations of land uses within clusters are measured. For the final phase, spatially distinct clusters are treated as individual landscapes. The remainder of the paper describes the approach and methods as they relate to the discussion presented above. The results of the analysis are not presented.

**Step 1: Selection of Parcels with Functionally Complementary Land Uses**

Candidate parcels for inclusion within clusters are defined on the basis of previous research as those containing medium-density residential uses, retail and service uses, and school sites. This preselection process addresses the issue of land use functional complementarity for, in the present case, travel behavior purposes. The range of preselected parcels is intentionally kept simple. Other land use types such as office uses can easily be added to the area within clusters.

The selection criteria for residential parcels are based on the distribution of net densities within residential land use classes. Included and coded are, first, all parcels with more than 25 dwelling units/ha, all parcels with attached or multiple housing units, and mobile home parks. Retail or service parcels are the second class of parcels and include convenience services likely to generate walking trips such as supermarkets, dry cleaners, restaurants, banks, and convenience stores and institutional uses such as post offices and libraries. The third class of parcels is school sites.

**Step 2: Delineating Clusters as Groups of Spatially Complementary Land Uses**

Parcel polygons are first converted to a raster- or grid-based data model to aid in spatial analysis with GIS software. Grid cells are set at 10 m/side, a size smaller than all but a few parcels, which minimizes spatial distortion from the original parcel polygons. Second, each grid cell is assigned one of the three land use categories discussed above. All other parcels in the data set are assigned a value of "no data." The grid data now no longer represent individual parcels but patches of homogeneous land uses. These patches become the spatial unit of analysis for cluster delineation. Figure 2 shows the four classes of patches and the tendency of the first three classes to agglomerate.

To group individual land use patches into clusters, the area of all three patches identified in the first three classes is extended in each direction by a specified distance. Any cell between patches located within twice this distance from each other is thus filled in. This fill area is termed "connective tissue" and is treated as a patch type. At this point in the delineation process, areas that are not yet filled but that are entirely surrounded by patches remain. These are islands within grouped patches that are filled by using grid analysis routines designed for hydrologic modeling. The final step in the delineation of clusters is to shrink the perimeter of grouped patches by the distance specified in the first step. The resulting clusters approximate "convex hulls," that is, spatial sets of patches defined by a shape of minimum perimeter (Figure 3).

As shown in Figures 4 and 5, specifying different distances between patches for cluster delineation results in different amounts of connective tissue and overall patch grouping. As the distance is increased, cluster size also increases, resulting in fewer clusters with more connective tissue. Different distances were tested between patches at 60-m increments (or 30 m plus 30 m from the edges of two patches under analysis), and a strong break in the amount of connective tissues was found when the distance specified between patches was changed from 240 to 300 m.

The 240-m increment was used because the analyses show strong patterns of spatial concentration for patches in the selected land uses at distances shorter than 240 m. In areas developed after World War II, the areas of the clusters created by this increment are generally less than 800 m in radius and thus correspond to areas with reasonable walking distances between land use classes. In older, urban areas, clusters are larger, but they generally correspond to neighborhoods known to supporting walking (5). In Figure 2, the three selected land use classes and the gray zone of connective tissues depict the results of the clustering process at this increment.

**Measuring the Clusters’ Land Use Composition and Configuration**

Once clusters are delineated, their land use composition and configuration can be quantified. For this purpose, a public domain software package called Fragstats was used (72). Three different levels of analysis were used: patches (corresponding to the individual zones of residential, retail-service, or educational land uses), classes (all patches classified with one of the three selected land use categories), and the landscape as a whole (in the present case, delineated clusters). Measures of composition and configuration complement traditional measures of land use such as density, in that they capture land use spatial functional complementarity and the grain of land use mixing. Also, because they are tied to parcel-level data, homogeneous patches avoid the modifiable area boundary problem. A large number of measures are available from Fragstats, but only a few are discussed here.

**Indices of Grain and Patch Shape**

A simple measure that begins to capture the grain of use mixing is mean patch size or patch density. A cluster with higher patch densities, that is, more patches per unit of area, implies more interspersion of uses than one with fewer, larger patches of use. Measures of the variations in patch size (for example, a standard deviation or coefficient of variation) also relate to the spatial distribution of different classes of patches. These measures may be most useful in conjunction with the measures of mix described below.

Two indices of patch shape are available in Fragstats, and both of these rely on ratios of patch area to perimeter. The first index compares the perimeter of a patch to an equivalent area configured as a circle (for vector data) or a square (for raster data). The second index uses fractal geometry and analyzes patches in terms of their fractal dimensions. These measures can be used to test the transportation implications for different configurations of land use patterns. For example, they can be used to test the travel implications for long, linear shopping streets versus those for areas with more compact, concentric retail patterns. Their predictive powers will need to be tested empirically.

**Interspersion, Juxtaposition, and Contagion:**
**Capturing Spatial Land Use Complementarity**

Fragstats contains several indices that measure the adjacencies between different land uses. They rely on the concept of unique types of edges between patches. For any pair of adjoining patches with
FIGURE 2  Clustering of selected land use patches within 240 m of each other in part of urbanized King County, Washington. Note the patterns of residential land surrounding retail-service land. Base data were the King County tax assessor's parcel files, 1998.
different land uses, a unique edge is created between them. Interspersion and juxtaposition are used together to describe the distribution of unique edges in the landscape. As the length of the different edge types around any given patch increases, the index value increases. A landscape with maximum interspersion and juxtaposition of patch classes is one in which the total amount of such edges is maximized.

The index will distinguish between two landscapes with the same total area and number of patches but in which one of the landscapes has patches of regular shape that minimize the length of shared boundaries and in which the other landscape has complex shapes that create long, sinuous boundaries. However, the index does not measure grain. It will not distinguish between landscapes that are similar except for the sizes of their patches. In other words, a landscape with uses in several large patches will have the same interspersion and juxtaposition as one with the same uses in many smaller patches (Figure 6).

The contagion index is similar to interspersion and juxtaposition, but it is also sensitive to the total length of the edge shared between classes as a percentage of the landscape as a whole. With all else being equal, the index will increase as the total length of the edge between patches in the landscape increases. Thus, the index will take into account the sizes of the patches and therefore the grain.

Unlike entropy and dissimilarity, these indices take into account the spatial distributions of different patches of land uses.

**Edge Contrast: Measuring Land Use Functional Complementarity**

The approach to identifying clusters described here is based on the preselection of patch types to control for the functional complementarity between land uses. However, land use functional complementarity can also be addressed by using the edge contrast index. The index weights different types of adjacencies between land uses. Landscape ecologists distinguish between high-contrast edges (for example, edges between grassland and forestland) and low-contrast edges (for example, edges between different types of forest). In a transportation context, adjacent patches of residential and retail use would create a high-contrast edge because these adjacencies are likely to induce travel. Low-contrast edges, on the other hand, would include those between educational and industrial uses, where significant travel is not expected. The relative weights that would be assigned to adjacent school and residential uses versus adjacent school and retail uses are more difficult to establish. In general, theory and empirical tests should be used to guide the researcher in developing weights. The index can be applied to individual patches, at the class level (for the mean contrast between a particular use and all others), or at the landscape level (for a measure of mean edge contrast). Note that at the patch and class levels the index captures spatial as well as functional complementarity. Because the index captures an aspect of land use that has not been well addressed previously, the assignment and testing of different weighting schemes is a potentially rewarding area for research.

**CONCLUSIONS**

The present work identifies two sets of tools and techniques that can be used to help better capture the characteristics of land use patterns that may affect transportation behavior. First, parcel-level GIS databases are now available in many metropolitan areas. They capture land use data at a scale that closely corresponds to land use patterns as they occur on the ground and offer more detailed, spatially referenced information about the built environment than previously used data offered. Second, measures developed by landscape ecologists to model patterns of land cover provide detailed and spatially explicit ways of measuring land use mix. These techniques permit researchers to consider land use functional and spatial complementarity as well as the grain or spatial distribution of heterogeneous land uses. The use of these techniques also requires researchers to be clear about the resolution of data, the scale of analysis, and the extent of the area across which analyses are conducted. Data resolution, analysis scale, and spatial extent can then be used as variables in themselves to probe the
FIGURE 4  Example of difference in size and number of clusters produced by using (a) 240-m versus (b) 480-m distance between patches to define groups.
The work presented here is at a beginning stage of modeling of land use composition and configuration reflected in development patterns. The next steps are to explore how the quantitative measures described are related to different development patterns. For given levels of land use intensity, for example, do measures of contagion vary systematically between the suburban clusters and areas developed before the 1930s that are known to generate walking trips? In which ways are these development patterns measurably different? Answers to these questions will allow planners to improve their understanding of the effect of the built environment on human behavior and the effects of policies on the built environment.

ACKNOWLEDGMENTS

The authors are grateful to the U.S. and Washington State Departments of Transportation for supporting this work through the Washington State Transportation Center (TRAC) and TransNow at the University of Washington. The authors also wish to thank the Puget Sound Regional Council for its support.

REFERENCES