PRELIMINARY DRAFT

Measuring and modeling urban sprawl: Data, scale and spatial dependencies

Elena G. Irwin*, Nancy E. Bockstael** and Hyun Jin Cho+

This draft: November 12, 2006

Invited paper prepared for the Urban Economics Sessions, 53rd Annual North American Regional Science Association Meetings of the Regional Science Association International, November 16-18, 2006, Toronto, Canada

* Corresponding author. Associate Professor, Department of Agricultural, Environmental and Development Economics, Ohio State University, irwin.78@osu.edu.

** Professor, Department of Agricultural and Resource Economics, University of Maryland, nbockstael@arec.umd.edu

+ PhD Student, Department of Agricultural, Environmental and Development Economics, Ohio State University, cho.228@osu.edu

Introduction

Despite a longstanding interest in the principle hypothesis of the monocentric model, that urban land use patterns are explained by distance from a city or suburban center, we know relatively little about fine-scale variations in urban land use patterns across an urban-rural gradient. While the theory has been well tested using population and employment density gradients, density as a descriptor of pattern omits many fine-scale attributes, including the degree of fragmentation, scatteredness and complexity that characterize the spatial arrangement of land in various uses. Fewer studies have sought to quantify variations in these fine-scale pattern characteristics and none have tested the extent to which the monocentric model may explain variations in these dimensions of pattern along an urban-rural gradient. This is striking given the historical importance of this model in explaining urban form, on the one hand, and the present day concerns over the relative amount of scatteredness or "sprawl" exhibited by urban land use patterns, on the other.

Data limitations are a primary reason for the lack of fine-scale, spatially explicit analysis of urban land use patterns. Such an analysis requires spatially explicit data on land use at a sufficiently disaggregate scale so that fine scale differences in urban pattern can be quantified. Appropriate land use data—data that are sufficiently fine scale and that accurately record the location and type of urban land use—are difficult to come by. Land use data are generated based on landscape images, most commonly aerial photographs or remotely sensed imagery, and thus the accuracy of the data depend on the reliability of the raw images to detect urban land use. As a result of different data generation methods, land use datasets can vary widely in their definitions, spatial resolution and overall accuracy, and not all datasets are appropriate for quantifying fine-scale variations in urban land use patterns.

This paper has two goals: first, to examine data and scale dependencies by comparing measures of spatial pattern across different datasets and across spatial scales of analysis; and second, to test the ability of the monocentric model to explain spatial variations in residential land use patterns across an urban-rural gradient. To accomplish the first goal, we compare the amount and pattern of urban land use as recorded by several different datasets for the state of Maryland, an urban-dominated state located along the eastern seaboard of the U.S. The datasets differ in the raw data and methods used to identify and classify land use as well as their spatial resolution. The differences that we find in the basic amount and pattern of urban land use as represented by these datasets are extreme. For example, 79 percent of the land in the state of Maryland that is classified as low density residential land use circa 2000 by the dataset derived from aerial photographs is classified as undeveloped by the dataset derived from remotely sensed images. In addition, we find striking differences in pattern measures across these different datasets and across different scales of analysis. Depending on the data, study region and scale of analysis, a researcher could logically deduce that sprawl is either rampant or nonexistent. Based on our analysis, we conclude that not all datasets are appropriate for quantifying sprawl and that the use of inappropriate data can easily lead to misleading conclusions regarding the incidence of sprawl and its underlying causes.

Secondly, we use these fine-scale pattern measures and the best data available to us to examine the extent to which accessibility to urban and suburban centers explain variations in fine-scale residential patterns across the state of Maryland while controlling for other sources of spatial heterogeneity. We perform this analysis at several spatial scales to examine substantive scale dependencies in the underlying processes that influence urban land use patterns. We find significant structural change in parameter values across the different scales of analysis. Our results suggest that the monocentric model provides a better explanation of residential sprawl patterns at a more aggregate scale of analysis than a local one and that even at a more aggregate scale, a substantial amount of the variation in pattern is not explained by this model. Lastly, we find substantial differences in this model's ability to explain observed variations in residential land use patterns in exurban versus urban and suburban regions of the state. While the model performs remarkably well when applied to residential patterns in urban and in particular suburban areas, it is exceedingly poor at explaining observed variations in pattern in exurban areas.

Measuring Land Use Patterns

Before proceeding, it is useful to define some terminology and concepts that distinguish types of spatial data and types of spatial scales, all of which are critical to the analysis of land use patterns. Two basic types of spatial data can be used to represent land use patterns: raster and vector data. Raster data divide the landscape into equal-sized cells so that each cell or pixel corresponds to a spatial location, where the cell value represents an attribute of the observation at that location (e.g., land use). On the other hand, vector data include points, lines and polygons that are geometrically defined and spatially referenced using coordinates that correspond to the points or the nodes that define the line or polygon. Irregularly shaped polygons are the most common vector representation of land use or land parcels. Either form of data may be used in pattern analysis and, as we discuss below, the spatial resolution at which these data are defined critically impact the measurement of land use pattern.

We distinguish among three different spatial scales in our subsequent discussion and analysis. These are the spatial resolution of the data, the scale at which pattern is quantified and the scale at which pattern is summarized:

Data resolution: The most disaggregate spatial data are those that are sufficiently fine scale to represent individual units, e.g., individual land parcels or houses. These objects are typically represented as vector data. Data defined at a more aggregate spatial scale abstract from the spatial delineation of individual objects and instead represent the summation or average value of these objects. For example, census block groups or tracts capture the aggregate number of residents within a spatially delineated area; land use polygons or cells capture the predominant land use or density and type of development within a given area. Both raster and vector data are subject to minimum data resolution constraints. Raster data are limited by the size of the cell used to represent the landscape attribute and vector data are limited by the minimum mapping unit used to delineate objects. For example, a minimum mapping unit of ten acres implies that the smallest polygon used to represent spatial pattern is ten acres and thus, smaller features of the landscape are not recorded. Clearly the scale of resolution matters greatly for representation of land use patterns. Calculating the amount of developed land using raster vs. vector data will produce estimates directly proportional to the size of the cell used. Assume that cells with built structures are coded as developed and those without structures are registered as nondeveloped. Then, in areas of low density development where houses are relatively far apart, a cell size of 60 square meters will produce an estimate of developed land four times that produced using the same underlying data but a cell size of 30 square meters.[1]

Pattern analysis: The spatial location of land uses relative to each other can be quantified using a variety of pattern measures, each of which is defined for a given spatial extent. Two basic approaches to measuring pattern are those that quantify the pattern surrounding individual units of observation (e.g., a land use cell or polygon) and those that quantify the shape of patches defined by homogeneous and contiguous units aggregated in space (e.g., a land use patch that is comprised of many contiguous and homogeneous land use cells or polygons). The former rely critically on the extent of the neighborhood surrounding individual units of observation; the latter rely critically on the extent of the area within which the patches are measured. An important distinction between these two approaches is that the individual-level measures generate as many observations on pattern as there are observations in the dataset. In contrast, a patch-level analysis often relies on a moving "analysis window" of a fixed size and therefore the number of observations is determined by the size of the analysis window relative to the size of the study region. Both are clearly dependent on the spatial resolution of the data and can give varying results, depending on the resolution. For example, using raster data with a resolution of 90 x 90 m^2 vs. 30 x 30 m^2 makes it more likely that small clusters of homogeneous land uses will be classified as isolated cells vs. multiple contiguous cells of the same land use type. Thus the same landscape represented with 90 vs. 30 square meter cells and quantified with an individual-level measure appears more scattered. On the other hand, aggregation from 30 x 30 m^2 to 90 x 90 m^2 cells will also cause the pattern at the patch level to become simplified with fewer edges and a higher area-to-perimeter ratio, signaling a more compact land use pattern.

Pattern summary: A third spatial scale is also possible. While it is certainly reasonable to describe the variation in pattern at the same spatial scale at which it is measured, it is sometimes useful to summarize the average pattern or variation in pattern at a higher spatial scale. For example, Burchfield et al. (2006) measure residential pattern using fine-scale data defined by a 30 x 30 m^2 cell size and summarize it at a more aggregate level by taking the average value across all residential cells within a metro region. Because pattern measures often exhibit substantial correlations and are not necessarily normally distributed, the aggregation results are often dependent on the scale at which the pattern is described. In addition, aggregating pattern to a regional level, e.g.,

[1] This will be true as long as the house situated in a 60mx60m cell is actually categorized as developed. Such a cell size is still less than an acre, but the proportion of the cell covered by driveway and house will be less than 10% and thus, depending on the land use classification rules, the cell may not be coded as developed.

multi-county or state, can mask important variations across or within counties that are critical to measuring and modeling residential land use patterns and their evolution over time.

Urban Pattern Analysis and the Monocentric Model

Economists have a long tradition of measuring variation in urban land use pattern using population and employment density gradients. The traditional approach assumes that the shape of the density gradient corresponds to a negative exponential function, which specifies the following relationship between density and population: log(population) = $\alpha + \beta$*distance + ε, where α is a constant, β is the estimated density gradient, and ε is a random error. Density is measured as population per square mile of land area and distance is measured as miles to the center of the business district. Census tract or blockgroup is the most common unit of observation, although a few studies have used counties or zipcodes. Using this type of model, a number of studies have estimated density gradients and found them to be downward sloping with distance and flattening substantially over time (e.g., Jordon, Ross, and Usowski, 1998; Mills, 1970). Other have employed flexible estimation techniques to allow for non-linearities or discontinuities in the density gradient that reflect the multicentric structure of modern cities (e.g., Alperovich, 1995; Brueckner, 1986; McMillen, 1994).

This approach to measuring urban land use patterns is useful for an aggregate analysis along an urban gradient, but has typically relied on population density at a census tract or even larger spatial scale and thus fails to capture the fine-scale characteristics of land use pattern that are often of interest from a more practical perspective. Local planners and policymakers are primarily concerned with the degree of contiguity and compactness of urban land use patterns, which have major implications for the cost of public services, traffic congestion and environmental impacts on wildlife habitat and water quality. Highly disaggregate, georeferenced data on land use, together with the power of Geographic Information Systems (GIS), have enabled researchers in recent years to begin quantifying fine-scale characteristics of urban land use patterns. Most of this work has been done by geographers, urban ecologists and others outside of economics. A popular approach that borrows techniques from landscape ecology (Turner, 1989) quantifies various aspects of urban land use patches, defined by contiguous observations (e.g., cells or parcels) with a homogeneous land use. As discussed above, these so-called landscape metrics useful for quantifying various aspects of land use pattern, including the fragmentation, complexity of shape and relative dispersion of urban patches within a given landscape (e.g., Sudhira et al. 2004; Seto and Fragkias 2005; Cutsinger et al. 2005). Other researchers have made use of spatial statistics, including measures of spatial correlation (e.g., Tsai 2005), to quantify the degree of sprawl in a metropolitan area. When applied over time, such analyses provide a description of how urban land use patterns have evolved within a given area. For example, Carrión-Flores and Irwin (2004) use several landscape metrics to quantify changes in the land use patterns within a representative Ohio exurban county between 1956 and 1996. The results demonstrate that, over this time period, a large increase in residential development has occurred, but, unlike commercial and industrial development, the pattern of residential land has become more fragmented rather than clustered over time.

Luck and Wu (2002) provide one of the few analyses of variation in fine-scale pattern across an urban gradient. Using fine-scale vector data that delineate individual land use parcels from the Phoenix metropolitan area, they calculate a series of landscape metrics along the urban gradient. They use a moving window to analyze variations in pattern along this gradient at two different scales of analysis: a 5 x 5 and 15 x 15 km^2. The results for the residential and urban land use patterns provide some support for a monocentric structure, but also indicate additional complexities. For example, the mean patch size of residential and urban land use patches increase with distance from the city center. However, they also find evidence that the pattern of land use near the city center is highly fragmented and more spatially complex than in outer lying areas. The authors conclude that, because separate concentric zones of land uses were not identified by the analysis, the results indicate that neither the monocentric model (referred to as the concentric ring model) or polycentric model (referred to as multiple nuclei theory) of urban form adequately explains the observed pattern. However, this conclusion is based solely on a descriptive analysis that fails to control for other sources of spatial heterogeneity that would naturally lead to some heterogeneity in pattern and thus does not provide a statistical test of the monocentric model. Nonetheless, their analysis suggests additional metrics that can be used to describe land use pattern along an urban gradient.

A recent cross-sectional analysis of metropolitan sprawl by Burchfield et al. (2006) is one of the most ambitious attempts to quantify the fine-scale spatial pattern of residential land use. Using datasets that correspond to circa 1976 and 1992 respectively, they measure the scatteredness of residential land use patterns across U.S. metropolitan areas by calculating the proportion of undeveloped land within a square kilometer of each 30 x 30 m^2 residential land cell located within a metropolitan area. This measure is then averaged across all residential cells in a metropolitan area to yield an index of sprawl for each metropolitan area. The authors then examine several hypotheses that seek to test the implications of the monocentric model for variation in residential land use patterns *across* metropolitan areas. Because they limit their analysis to a metropolitan scale, they do not examine the primary hypothesis of the monocentric model—that distance from urban centers explains variations in residential land use patterns. A surprising conclusion of the paper is that, while substantial variation in residential sprawl exists across U.S. metropolitan areas, the relative amount of residential sprawl has remained constant over recent decades. However, this conclusion is based on a comparison of the residential sprawl measure applied to two different datasets, one based on aerial photography and the other based on remotely sensed imagery. As we demonstrate in the following section, remotely sensed data are not appropriate for measuring low density residential land use and thus the authors' conclusions are likely to be biased by data limitations.

Data and Scale Dependencies in Measuring Sprawl

Our principal motivation in this first part of the paper is to better understand the limitations of available land use datasets for measuring urban sprawl and the potential pitfalls of relying on inappropriate data or a single spatial scale in attempting to draw conclusions about the incidence of sprawl. We take as our starting point the recent analysis by Burchfield et al. (2006) (hereafter, BOPT). There are many innovative aspects of this paper, including their intuitive measure of residential sprawl, defined as the proportion of undeveloped land within a given neighborhood of a residential land use cell. A shortcoming of many of the studies that use landscape metrics to gauge sprawl is that they often fail to provide any intuition for how these measures should be interpreted as measures of sprawl. By using a straightforward measure of residential scatteredness, BOPT avoid this pitfall. In addition, there are many aspects of this paper with which we agree, including the assertion that previous datasets have not been well suited to test hypotheses about these patterns. In particular, the National Resource Inventory (NRI) data is simply too aggregate to be used in any meaningful way to measure sprawl.

However, there are several key aspects of the BOPT analysis with which we disagree. First, we question the validity of directly comparing spatial data derived from different data generation processes to determine the location and pattern of new development. Land cover and use data have traditionally been generated for specific purposes and thus the methods and resulting data are not designed to be consistent with different land cover/use datasets (Homer et al. 2004). Depending on the data generating process used, e.g., the spatial resolution, spatial accuracy and the rules used to define various land use categories, data may or may not be comparable. We believe that it is this consideration that has kept geographers from performing the sort of direct spatial comparison of two different land use/cover datasets that BOPT rely on to derive their measure of new development. A publication by U.S. Geological Survey researchers that precedes the BOPT paper by several years contains a series of maps using the same two datasets that BOPT use to visually represent urban patterns at two different time periods for selected metropolitan areas, but stops short of spatially overlaying these two datasets to derive estimates of spatially explicit land use change (Auch et al., 2004). The fact that this group of researchers (which includes some of the very same researchers who generated these datasets in the first place) refrained from any direct comparison of these two datasets is suggestive of the limitations inherent in such an analysis.

Second, remotely sensed data detects land cover rather than land use. This poses a problem when attempting to measure the extent of low density development, frequently defined as development on lots of at least one acre but often extending to densities as low as one house per five or ten acres. Given the fine scale at which the NLCD data are recorded (a 30 x 30 m^2 cell size, which equals about 0.22 acres), a house on a five acre lot would logically be categorized by satellite data as one cell in development and approximately 22 undeveloped cells. A land *use* data set, however, would assign all 23 cells to low density development. Whether development should be defined simply as the cell upon which a structure resides or whether it should include the entire housing lot leads to substantial confusion in comparing measures across different types of data sets.

Given that a majority of new development in the U.S. is estimated to be on relatively large lots,[2] this inconsistency will lead to considerable discrepancies in measuring residential development.

The distinction between how much land is in developed use versus how scattered the structures are will plague any analysis that attempts to measure sprawl. It will also confuse any measure of the *amount* of urban land. Three hundred square meters is a generous allotment for the footprint of a house and its accompanying driveway – about a third of the size of the 30 x 30 m^2 cells used in the BOPT analysis to detect new development. But a one acre lot is over 4,046 square meters or about 4.5 of these cells. Counting cells instead of lots would underestimate developed land of this density by a factor of four. And of course, the larger the lot size, the greater the underestimate.

Third, the problem with using remotely sensed data extends beyond the discrepancies between over land use and land cover data sets. It is unclear whether the data used by BOPT are appropriate for measuring urban sprawl, largely because of the difficulties involved in detecting low density residential development using remotely sensed images. According to information posted on the USGS website (http://landcover.usgs.gov/ftpdownload.php), the 1992 data are limited in part by a relatively small number of aerial photographs used for ground truthing, which make reliability of the data for areas smaller than a state uncertain. Users are cautioned to ensure that the data provide sufficient reliability before attempting to use the dataset for local (i.e., sub-state level) analyses. According to McCauley and Goetz (2003), it is difficult to discriminate low-density residential areas due to the range of land cover types within this specific land use and their associated spatial variability. Thus a spectral "signature" for low density development cannot be reliably discerned and these data will undercount low density development due to classification errors.

Lastly, we question whether a single statistic applied at a single scale of analysis is adequate to capture the potential complexities that may arise in measuring and modeling sprawl. A common feature of many spatial patterns is scale dependence, an issue that geographers have long grappled. Such dependence may be spurious or substantive--e.g., a pattern may exhibit strong positive spatial correlation at an aggregate scale, but substantial dispersion at a more disaggregate scale. In either case, an analysis that is limited to one spatial scale fails to account for such scale dependent effects. Another concern relates to the perhaps unintended consequences of measuring pattern relative to each individual residential cell. As with any individual-level pattern measure, the results are weighted by the total amount of cells in a particular land use. In this case, because remotely sensed data miss much of the low density development, the residential pattern associated with low density development is underweighted.

Data Description

[2] Analysis by researchers at the Economic Research Service (U.S. Department of Agriculture) based on the U.S. Census American Housing Survey data reveals that, between 1994 and 1997, lots of one acre or greater accounted for over 90 percent of new land converted for housing (Heimlich and Anderson, 2001).

Figure 1 provides a base map of our study region, the State of Maryland. With an estimated population of 5.6 million in 2005, Maryland ranks as the 19th largest state in terms of population, but with just over 12,400 square miles is the 42nd largest state in terms of land area. The 95 corridor connecting Philadelphia, Baltimore and Washington, DC passes through the center of the state and many of its counties contain suburbs or the latter two major cities. In addition to Baltimore, the state includes a number of suburban cities and smaller metropolitan areas with sizeable populations, as illustrated in Figure 1. The western part of the state, as well as sections of the Eastern Shore of the Chesapeake Bay, are largely rural, and even the more developed counties still retain large tracts of undeveloped land. Nonetheless, given its relatively small size and large urban population, Maryland is a much more urban state than the average U.S. state.

To perform the comparative analysis of data and scale dependencies, we make use of several different datasets.[3] These are:

1. **1973 and 2000 Maryland Department of Planning (MDP) land use data:** These are vector data derived from high altitude aerial photography and ancillary data. Both datasets have a minimum mapping unit of ten acres. However, the 2000 urban land use data were further refined using parcel data from Maryland Property View data, which provide these land use data with greater spatial accuracy. Both datasets distinguish among eight different urban land use categories, including low, medium and high density residential, commercial, industrial, institutional and urban open space. Residential density corresponds to the following number of dwelling units per acre: 0.2 – 2, low density; 2-8, medium density; greater than 8 and multi-family, high density.

2. **2001 National Land Cover Dataset (NLCD):** These are raster data derived from remotely sensed Landsat satellite TM imagery that has a spatial resolution of 30 x 30 square meters and that correspond to the late 1990s and early 2000s (circa 2001). Land use classifications were performed with a so-called decision tree classifier (Homer et al., 2004), an algorithm that provides a means of assigning land cover classes based on combining spectral and ancillary data to analyze and associate spatial variability with a particular land cover class. The generation of these data are said to be a substantial improvement over the NLCD 1992 data (Homer et al. 2004). Urban land is classified into developed open space (includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses; impervious surfaces account for less than 20 percent of total cover); low intensity developed (impervious surfaces account for 20-49 percent of total cover); medium intensity developed (impervious surfaces account for 50-79 percent of the total cover) and high intensity developed. These data are available for almost all of Maryland with the exception of the most western portion of the state, which is comprised of two small metro counties (Washington and Allegany), one non-metro county (Garrett) and the western portion of a suburban county (Frederick).

[3] Details regarding the data generation and land use classifications for each dataset can be found in Appendix A.

3. **2000 Maryland Department of Planning Maryland Property View data (MPV):** These are geocoded centroids of each parcel in each available county, with attached parcel attributes including whether a structure is found on the property and the date the structure was built. We have constructed a GIS coverage of the points which mark the centroids of parcels with structures on them. This provides an approximation of the location of the actual structures. These data are available to us up to a year built data of 2001 for a thirteen county central Maryland region (this region includes Baltimore City and the area around Washington D.C.; it extends from Frederick in the west to Cecil in the east and from Carroll and Baltimore in the north to St. Mary's in the south). We are also able to use the year built data to recreate historical patterns of residential housing structures.

4. **2002 Howard County parcel boundary data (HPB):** These are vector data generated by geocoded counties tax maps from the beginning of 2002, which delineate the boundaries of privately and publicly owned parcels. The data include attributes of the parcels, including the land use category assigned by the tax assessor and an indicator variable as to whether or not a structure is on the parcel. We use these data to assign one of the following land uses to each parcel: single-family residential, multi-family residential, non-residential structure, developed open space, undeveloped and protected undeveloped. We further specified a ten acre maximum cut-off for land parcels classified as single-family residential structures. This avoids counting rural residential land (e.g., a farmhouse on 100 acres) as being developed.

5. **2002 Howard County building footprint data (HBF):** These are vector data of the footprint of every structure in Howard County. We use these data to construct a raster dataset that correspond to the location of residential and non-residential structures in Howard County.

Fine-Scale Land Use Comparisons

Because we have the most spatially accurate data for Howard County (both parcel boundary data and building footprints), we focus first on this county. Howard County is a suburban county north of Washington D.C. and west of Baltimore (see Map 1) that contains a noticeable urban-rural gradient. In Map 2 we compare for Howard County the visual resolution and pattern of land use represented by each of three datasets circa 2000: the 2002 Howard County Parcel Boundary data (hereafter, HPB), the 2000 Maryland Department of Planning data (hereafter, MDP) data and the 2001 National Land Cover Dataset (hereafter, NLCD). It is clear that these data vary substantially in their spatial resolution and identification of land use and thus exhibit noticeable differences in their representation of urban land use pattern. Maps 3 and 4 provide closer views of a high residential density area in the eastern portion of the county and a low density residential area in the western portion of the county.[4] These maps illustrate the HPB parcel boundary and land use data (Maps 3c and 4c) and their overlay with MDP (Maps 3a and 4a) and NLCD (Maps 3b and 4b) land use data. While the MDP data appear to do reasonably well at approximating the land use parcels in both the high and low density areas, the NLCD data clearly do not. The NLCD data record much of the high density development (Map 3), but record almost nothing of the extent and pattern of low density development (Map 4).

To systematically explore differences across these datasets we first compare the total land area by aggregated land use categories as recorded by the HPB, MDP and NLCD datasets for this single county (Table 1). The HPB data record just over 62,500 acres (42 percent of the total land area) as developed, compared to about 57,000 acres (35 percent) according to MDP and either 11,000 or 32,000 acres (7 or 19 percent) according to NLCD depending on whether cells classified as developed open space as counted as development or not. Thus we find that, while the MDP data slightly undercount the amount of developed land in the county, the NLCD data record at best only *half* the total amount of developed land in the county.

We further analyze these differences across datasets using an area-to-area comparison that applies an identically defined grid as that of the NLCD data (i.e., a grid with the same origin coordinates, number of rows and columns, and, in this case, a cell size of 30 x 30 m^2) to the two vector representations of land use (HPB and MDP). Thus it provides an explicit spatial comparison of how each dataset records the land use associated with a specific location. The results, illustrated in Table 2, compare how both the MDP and NLCD datasets compare to the HPB land use data. The most striking differences are in the classification of single-family residential land. The MDP data count 66 percent of this land as in either a low or medium residential use and 32 percent as undeveloped. The NLCD data classify just 22 percent of this land as being developed (including 15 percent classified as developed open space) and count 75 percent as being undeveloped. Both datasets classify a substantial amount of non-residential development as undeveloped (46 and 48 percent for MDP and NLCD respectively) and both are very accurate on the classification of undeveloped land (87 and 89 percent for MDP and NLCD respectively).

Such a comparison may be misleading, however, to the extent that the remotely sensed NLCD data are unable to identify the entire developed land parcel. For example, a cell is classified as low intensity developed only if 20 percent of the cell is found to be in structures or other impervious surface. While the 'developed open space' category may pick up some low density development, the dataset clearly misses the vast majority of this type of development. In contrast, the HPB data record the full extent of any residential lot (up to 10 acres) as developed.

[4] These views were chosen arbitrary based on their contrasting residential densities and are each typical of how high vs. low density development is recorded by each of the datasets.

A fairer test, then, is to examine whether the NLCD data at least pick up the built structures in low density areas. Focusing on residential housing, because that is the essence of sprawl, we use Howard County's building footprint data to calculate the percent of actual houses that are registered as any developed cell in the NLCD data.[5] To do so, a grid of the Howard County housing footprints (for lots of 10 acres or less) is overlaid on the NLCD grid. A house is considered to be properly recorded as a developed cell if the cell the house falls on is categorized as developed in the NLCD dataset or if an adjacent cell is categorized as developed. This attribution scheme allows for slight shifts in the geo-registering of the two data sets and also handles cases where houses straddle cell lines. Even with this rather generous attribution scheme, over 30% of the houses that were present in Howard County as of 1992 are missed by the NLCD data. The failure of NLCD data to reflect the extent of residential development in Howard County is not random. A probit analysis, attempting to explain the observed vs. missed houses using a) distance from Route 95 (the center of Howard County development in the eastern part of the county) and b) size of the lot, is convincing. The ratio of the estimated coefficient to standard error exceeds 50 for both variables. Thus, the ability of the NLCD data to properly record development declines dramatically with housing density and also with distance from the highly developed eastern part of the county.

In sum, we find that while remotely sensed data regularly classify higher density developed land as developed, these data systematically misclassify the vast majority of isolated houses that make up low density sprawl. The discrepancy goes beyond definitional differences in what constitutes developed land and whether low density development corresponds to the land parcel boundary or just the built portions. The remotely sensed data miss substantial portions of low density development, including the built structures, and appear to be seriously biased towards misclassifying development that is not surrounded closely by other development. These results cast doubt on the validity of using such data to conclude anything about residential scatteredness or sprawl.

Parcel boundary data is, unfortunately, not available for the entire state of Maryland and limiting our analysis to Howard County would be overly restrictive. In order to extend our analysis to a larger region we rely on a comparison of the MDP and NLCD data, which allows us to investigate a number of different issues relating to data and scale dependencies. Based on our previous comparisons of these data to the HPB data for Howard County, we conclude that, although the MDP data have a coarser spatial scale of resolution, these data are more accurate in recording the location and pattern of residential land use and in particular, low density residential use.

We examine the amount of state land in each broad category of use according to the MDP and NLCD datasets (Table 3). The differences between the datasets across this broader region level mirror the differences found in the Howard County analysis. While the NLCD data report 12.8 percent of the total land area as developed, the MDP data

[5] For this experiment we compare the 1992 NLCD data, as used by BOPT, with the houses built as of 1992 in the Howard County data set. A similar comparison could be made using the 2001 NLCD data, but has not yet been completed.

identify a total of 20.8 percent in development. An area-to-area comparison (Table 4) reveals a similar striking difference in how low density development is coded. A full 79 percent of the land designated as low density residential land use by the MDP data are coded as undeveloped by the NLCD data. Another 12 percent is coded as developed open space. In contrast, 26 percent of medium density and 19 percent of high density residential land is classified as undeveloped. As with the Howard County comparisons, both datasets largely agree on the designation of undeveloped land: 90 percent of land classified as undeveloped by MDP is also classified as undeveloped by NLCD. The similarity of the comparisons between the MDP and NLCD data evaluated for Howard County vs. the Maryland region suggest that the same limitations of the NLCD data found for Howard County are likely to apply to the larger regions as well.

Residential Land Use Pattern Comparisons

While the datasets reveal substantial differences in the amount and location of land categorized in different uses, it is possible that the qualitative aspects of the land use pattern are nonetheless similar. To examine whether data dependencies extend to the qualitative aspects of pattern analysis, we compare the spatial pattern of residential land use that is represented by these datasets by applying the same measure of residential sprawl used by BOPT. This statistic measures the proportion of undeveloped land within a given neighborhood of a residential cell and thus varies from zero to one. A higher number means that a larger proportion of the cells within the neighborhood of a developed cell are undeveloped cells, suggesting greater scatteredness of the residential pattern. We then examine the average value and frequency distribution of each and, to examine variations in the pattern measure across various spatial scales, we also calculate the statistic for varying neighborhood sizes. Square neighborhoods are defined relative to individual cells so that each neighborhood is centered around each individual residential cell.[6] The central questions that we ask are (1) does this measure of pattern differ across the different datasets and if so, how and (2) is the pattern measure scale dependent?

We first compare the MDP and NLCD data and a third dataset, Maryland Property View (MPV) 2001 data, which approximate the location of residential houses using parcel centroids based on geocoded tax assessment data. Because the MPV data are available to us only for a thirteen county central Maryland region, we limit the MDP and NLCD measures to this region as well for comparability. The comparison of pattern across these three datasets allows us to compare three distinct datasets with varying definitions of residential land, spatial resolution of the data and data generation methods. Because the MDP data are defined at the lowest (i.e., most aggregate) spatial resolution, we expect these data to exhibit a more clustered, biasing the pattern measure for these data towards zero. However, the MDP are also reasonably good at tracking the pattern of low density development and thus, to the extent that this pattern is scattered, this will offset the downward bias. On the other hand, the MPV data are parcel centroids and as such do not represent land area; instead, they provide a reasonable approximation for the location of residential structures and can be used to quantify the pattern of residential

[6] This relative neighborhood measure differs from the BOPT analysis, in which neighborhood boundaries were non-overlapping and fixed in space rather than centered on each residential cell.

housing vs. land. For this reason, we expect that these data will exhibit a high degree of scatteredness. Lastly, because NLCD data are better at tracking structures vs. developed land, we expect that the NLCD pattern measure will fall closer to the MPV endpoint. However, if NLCD misses a substantial proportion of the more isolated structures and scattered development, this will bias the measure downward and cause it to fall closer to the MDP measure.

To make the pattern measures comparable, the MDP and MPV data are converted to a raster format using a 30 x 30 m^2 cell size. Doing so with the MPV data requires an implicit assumption about the average footprint of a residential structure. A 30 x 30 m^2 cell size corresponds roughly to a 0.22 acre lot. Given that the low intensity developed category requires a minimum of 20 percent impervious surface, this assumption translates roughly into an average footprint of 1900 square feet. The BOPT measure is then calculated for each residential cell in each dataset at four different spatial scales: 0.5, 1, 3 and 5 square kilometer neighborhoods. The MDP data identify three residential land use categories (low, medium and high density residential) and we consider all three in the current analysis reported here. The NLCD data do not exclusively identify a residential land use category; however, according to the data documentation, the developed low and medium intensity categories most commonly include single-family residential housing and therefore we use these two categories to measure residential land.[7]

Figure 1a-d illustrate a series of histograms that describe the frequency distribution of the BOPT residential sprawl measure for each of the three datasets calculated at varying neighborhood extents. Table 5 lists the average values associated with each histogram. A number of interesting results emerge. It is immediately obvious that the pattern, as represented by the sprawl measure's frequency distribution from zero to one, varies substantially across datasets. As expected, the MPV pattern exhibits the most scatteredness, with an average value ranging from 63.8 to 83.9. With the exception of very high values, the MDP distribution lies everywhere above the other two distributions, reflecting the fact that this dataset records a greater total amount of residential land. For the two smaller neighborhoods, 0.5 and 1 km^2, the MDP measure has a large number of zero values, reflecting a more clustered pattern at these spatial scales. The NLCD data also record a substantial number of zero values at these smaller scales and even outnumber the MDP data at zero at the 3 and 5 km^2 scales. This is striking given our expectation that the NLCD data would reveal a less clustered pattern due to their finer scale resolution. Both the MDP and NLCD data exhibit very clear scale dependence: whereas both are monotonically decreasing when measured with at the smaller neighborhood scales,[8] this is no longer true at the larger scales of analysis. In contrast, the MPV pattern measure appears to be relatively invariant to neighborhood scale.

---

[7] Based on our preceding analysis, we have found that the 'developed open space' category contains a fair amount of residential development and therefore we are also conducting the comparison with this category included, but have not yet completed this analysis.

[8] This is true with the exception of the regular 'blips' in the distribution, which are the most noticeable at the 0.5 km^2 scale. These are caused by the fixed cell size which, at smaller neighborhood sizes, causes the sprawl statistic to not be continuously distributed. This artifact disappears at larger neighborhood sizes because there are a greater number of cells and therefore a more continuous distribution of area proportion.

The results confirm that the MDP data exhibit a substantially more clustered pattern than the MPV data, as expected. However, the NLCD data are also substantially more clustered than the MPV data, which is unexpected. This is most likely because the NLCD low and medium intensity developed categories tend to pick up higher density, more clustered development and miss the low density residential development.

Next we perform the same analysis using just the MDP and NLCD data for the larger Maryland region. Figure 2a-d reveals very little difference in the overall distribution of pattern between this larger region and the smaller central region (Figure 1). It is noteworthy that the distributions are qualitatively more similar at smaller scales than the large scales. Particularly when evaluated at the 5 km^2 scale, we find that the MDP data exhibit an increase in the relative amount of scattered development whereas the NLCD data continue to show a relative decline in the number of highly scattered residential cells. This is reflected in a divergence of their average values, which are relatively close when evaluated at the 0.5 km^2 neighborhood size (30.7 and 27.8 for MDP and NLCD respectively) and diverge at larger scales (e.g., 54.6 and 43.2 at the 5 km^2 scale for MDP and NLCD respectively).

Variations across an Urban-Rural Continuum

To capture differences in residential pattern across an urban-rural continuum, we plot these two histograms by type of county. Using a modified version of the urban-rural continuum codes used by the Economic Research Service of USDA, we classify counties as falling into one of five categories according to their total population level and their designation as belonging to a metropolitan, micropolitan or rural region (see Appendix B for definitions and a map of counties by urban-rural type). The result is 21 percent classified as large urban; 12.5 percent as suburban; 25 percent as exurban; 16.5 percent as small metro/micro and the remaining 25 percent as rural. Table 6 illustrates the distribution of developed land uses according to the MDP data across this continuum. Not surprisingly, the majority of residential development and non-residential development is located in the large urban counties. However, suburban and exurban counties also contain substantial amounts of low and medium density residential development. The pattern distribution by these county types for both the MDP and NLCD data are illustrated in Figure 3a-d for the 1 km^2 neighborhood scale and in Figure 4a-d for the 5 km^2 neighborhood scale and Table 7 contains the average sprawl value for each dataset and neighborhood scale. The results show wide variation in the pattern distribution across the urban-rural gradient and once again, across datasets and neighborhood scale. The large urban counties exhibit the least amount of scatteredness, as expected. The pattern in the small urban, suburban and exurban counties exhibits a fair amount of scattered development. Interestingly, the distribution for the NLCD data in the small urban and exurban areas is relatively flat whereas the MDP data display a clear increase in the distribution for both regions and at both scales of analysis from a value just greater than zero to around 60 or 65 for the 1 km^2 scale and around 80 for the 5 km^2 scale. These areas and in particularly exurban areas, contain a higher proportion of low density, scattered development; this is reflected by the increase in the frequency distribution for the MDP data across a large portion of the range of the data. The relatively flatness of the NLCD distribution is explained by the fact that these data do not capture the vast majority of the more scattered, low density exurban development in these less urbanized regions.

Based on the pattern analyses illustrated in Figures 1-4 and Tables 5-6, we conclude that there is clear data dependence in this measurement of residential sprawl patterns. In both cases, there is clear scale dependence as exhibited by the shifting frequency distributions across the varying neighborhoods. Conclusions regarding the incidence of sprawl, however, would vary substantially depending on the dataset. Using the NLCD data, one would conclude that the relative proportion of residential sprawl to more compact development is not high and that exurban areas tend to have an equal proportion of more compact and more scattered development. On the other hand, conclusions based on the MDP data would lead to substantially different conclusions regarding the incidence of sprawl. Here we conclude that the proportion of sprawl increases with increasing neighborhood extent and that particularly in exurban regions, the relative amount of highly scattered residential development is high.

Variations in Residential Sprawl across Time

Lastly, we examine variations in pattern across time according to a comparison of the 1973 and 2000 MDP datasets. The MDP data are particularly useful in this regard as the 2000 data were developed with the intent of making them as consistent as possible across time with the 1973 data. Figure 5 illustrates the results of the MDP comparison between the 1973 and 2000 datasets evaluated at two neighborhood scales (1 and 5 km^2). The distribution for both the frequency count (histogram) and relative proportion (in a sense, the probability density function) is given so that both the absolute and relative differences in pattern are made clear. Interestingly, the relative pattern evaluated at the 1 km^2 neighborhood shows remarkably little difference across time. This neighborhood extent is the same of that used by BOPT, who also find the distribution in this measure of sprawl for all U.S. metro areas to be relatively constant between 1976 and 1992. More differences are discernable in the relative pattern evaluated at the 5 km^2 scale; in particular, the relatively large number of extremely scattered residential locations is reduced and the relatively number of locations with substantial, but not extreme scatteredness (e.g., over 90), is larger. Despite these discernable differences in pattern, the relative stability of the pattern over time is noteworthy.

A different picture merges, however, when we examine the differences in pattern by county type. Figures 6 and 7 illustrate these comparisons for the 1 and 5 km^2 neighborhoods respectively. There are more substantial shifts in the amount and distribution of residential scatteredness in suburban and exurban areas than what is represented by the aggregate plots in Figure 5. In both suburban and exurban counties, the pattern shifts away from a pattern skewed towards extreme scatteredness (around 80 for the 1 km^2 neighborhood and over 90 for the 5 km^2 neighborhood) towards a pattern that is more equally distributed across moderate and moderately high levels of scatteredness (e.g., 40 to 60 and 50 to 80 for the 1 km^2 and 5 km^2 neighborhoods respectively). In contrast, the relative pattern in the large urban counties remains virtually unchanged during this time period.

These results demonstrate the limitations of summarizing pattern by aggregating an individual-level measure. In this case, summarizing the pattern at the scale of a multi-county region masks important differences in the qualitative pattern of residential land use across counties. This problem is due to the inherent bias of the pattern measure, which is evaluated relative to individual residential cells and thus is proportionate to the number of residential cells. Such an approach biases the measure towards the pattern exhibited by regions that contain the most residential cells, namely larger urban areas. In this case, the observations from the 2000 data that are from large urban counties comprises 43 percent of the total number of residential cells vs. 14 percent for suburban counties and 21 percent for exurban counties.[9] In comparing the aggregate pattern in Figures 5 with the disaggregated patterns across county types in Figure 6, the bias in the aggregate pattern from large urban counties is clear. As a result, important changes in the relative pattern of residential scatteredness in suburban and exurban areas are masked when only the aggregate pattern for the region is considered.

Finally, we note that the interpretation of this measure of residential scatteredness as "sprawl" is itself scale dependent. Sprawl is an elusive concept and we make no attempt to define it here. Rather, we simply point out that the degree to which this measure tracks our visual sense of sprawl is dependent on the neighborhood scale at which the measure is defined. For example, a person standing at the boundary of two 0.5 km^2 neighborhoods with residential development at their center would be in the middle of residential development spaced just over a tenth of a mile apart and in this case, may very well conclude that the pattern is one of sprawl. However, the same person standing at the boundary of two 5 km^2 neighborhoods with residential development at their center is 1.5 miles away from either residential development. It would be difficult to label such isolated development as "sprawl." Instead, a sprawl pattern is more likely to be discerned as these neighborhoods start to infill with hop-scotch development, i.e., as the measure of scatteredness *declines* in these neighborhoods. Viewed in this light, our results from the 5 km^2 neighborhood analysis suggest substantial increases in the incidence of sprawl over time in suburban and exurban areas.

Explaining Spatial Variations in Residential Sprawl across an Urban Gradient

Up to this point we have concentrated on the differences in measurement of land use patterns across datasets and varying scale of analysis. We now turn our attention away from data comparisons to explaining the observed spatial variations in the pattern of residential land use as represented by a single dataset. Given the superior ability of the Maryland Department of Planning (MDP) data in recording low density residential development, we rely on these data to represent the spatial location and pattern of land use. Our primary goal is to test the basic hypothesis of the monocentric model, that urban

[9] The measures from 1973 are similarly biased towards large urban counties with 59 percent of the total observations from these counties and just 11 and 15 percent from suburban and exurban counties respectively.

land use patterns are a function of distance from urban centers, and to examine the extent to which such a hypothesis is scale dependent. By 'scale dependent' we mean that the estimated parameters of the regression model vary significantly across different spatial scales of analysis. In particular, we hypothesize that the monocentric model may provide a better explanation of residential pattern at more aggregate scales of analysis. In addition to modeling the existing pattern of residential development in 2000, we also consider the historical pattern of residential development in 1973, which allows us to examine structural change across time.

We begin with the same measure of residential sprawl as above, the BOPT measure in which the proportion of undeveloped land around a residential cell is calculated for each residential cell in the dataset. For each observation in the 2000 and 1973 datasets, we use GIS to calculate a set of explanatory variables. These include the basic monocentric model variables—measures of accessibility to the two major urban centers, Baltimore and Washington D.C. We measure accessibility as the shortest distance via the major roads network from each residential cell to the nearest point on the beltway[10] surrounding each of the two major urban centers. We also measure distance via the major roads network to the nearest other major employment center, defined as the nearest other city, and distance to the nearest place with a minimum population of 10,000 in 2000. Lastly, we account for exogenous sources of spatial heterogeneity by using a set of variables derived from the location's soil characteristics and topography. These measures include a series of indicator variables that are assigned a value of one if the location is particularly good for development (i.e., well draining soils and relatively flat slopes); bad for septic tanks (i.e., poorly draining soils); prone to flooding; and particularly good soils for farming.

Three nested models are estimated to investigate the relative influence of urban center proximity: the first includes measures of accessibility to the two major urban centers and the other two include controls for additional exogenous variations. To examine scale dependence, we calculated the sprawl measure and the set of explanatory variables for three different cell sizes: 30, 100 and 500 meters. In preliminary analysis, however, we found little difference in the regression results across these three different cell sizes and thus focus our discussion here on the results generated with the 100 meter cell size.[11] In addition, we calculate the residential sprawl measure using different neighborhood extents to examine whether results are dependence on this type of scaling effect. We find that the results are scale dependent in this respect and that the overall fit of the model improves with increasing neighborhood size. Because the results show a consistent pattern across the scale of analysis, we report the results for the smallest (0.5 square kilometers) and largest (5 kilometers) neighborhoods only. Lastly, to address the issue of an inherent bias in the underweighting of observations on pattern from lesser

[10] Distance for residential cells located within the Baltimore and Washington D.C. beltways are set to zero.
[11] Recall that we convert the MDP vector data into raster format by overlaying a grid with an assigned cell size. This operation assigns each cell the value of the majority land use that falls within the cell. Because the original data are in vector format, we are not constrained in the cell size that we use to transform the data.

developed areas, we also consider separate regressions for large urban, suburban and exurban counties.

Regression Results

Table 7 reports the results of the first model, which is simply an intercept term and the two accessibility measures for Baltimore and Washington D.C. For this and all subsequent models, we find that the linear-log model specification has a much better overall fit than a linear-linear model or a model in which the distance variables are allowed to vary nonlinearly, suggesting that the influence of proximity to urban centers declines monotonically with distance. The results of four different model estimations—the 1 and 5 km^2 neighborhood analyses for both 1973 and 2000—are reported. Considering first the results for the most recent dataset, 2000, we find that the total amount of variation explained by distance to the two large urban centers varies from 0.142 to 0.34 depending on the scale of analysis (as defined by the 1 or 5 km^2 neighborhood size). This result is consistent across time. The basic monocentric model explains a greater amount of the variation in the larger scale pattern of residential sprawl in 1973 vs. 2000. In comparing the results across 1973 and 2000, we find that the total amount of variation explained by the basic monocentric model is remarkably similar for the smaller scale analysis (with R^2 values of 0.136 and 0.142 respectively). Interestingly, the basic model explains a greater amount of the variation in the larger scale pattern from 2000 vs. 1973 (with R^2 values of 0.245 and 0.34 respectively).

Table 8 reports the results for the second model, which includes distance via the major roads network to the nearest other major employment center, defined as the nearest other city, and distance to the nearest place with a minimum population of 10,000 in 2000. These two additional distance measures are highly statistically significant in three of the four models—only in the 1973 model estimated with the larger neighborhood extent is the distance to nearest other city not significant. However, the addition of these variables adds very little to the overall explanatory power of the model. In comparing the R^2 values of each of the models with their counterpart in Table 10, we find that the overall fit of the model is essentially the same as that of the basic monocentric model.

The results of the third model, in which local sources of spatial heterogeneity are controlled, are reported in Table 9. This model accounts for local exogenous variations that make the location more or less suitable for development. All of these variables are significant and all of them are found to contribute positively to residential scatteredness with the exception of the flood indicator variable. We find some differences in goodness of fit across the two neighborhood scales. For the smaller neighborhood scale, the addition of these controls increases the overall model fit by only a very slight amount (from 0.14 to 0.165 for 1973 and from 0.14 to 0.157 for 2000). On the other hand, the overall fit of the model estimated at a larger scale is increased somewhat more (from 0.245 to 0.303 for 1973 and from 0.348 to 0.382 for 2000).

Table 10 reports a series of t-tests that examine whether the coefficients associated with the distance measures are significantly different across time (i.e., across the 1973 and 2000 datasets) and across the two neighborhood scales of analysis. The results show some structural change in the estimates across time, although the magnitude of the change is relatively modest in most cases. The Baltimore coefficient is not significantly different across the two time periods and the Washington D.C. coefficient increases significantly between 1973 to 2000. In terms of marginal effects[12], this translates into a relatively modest increase however; for example, a marginal increase in distance from Washington D.C. by one kilometer is estimated to increase residential scatteredness by 2.06 in 1973 and by 2.09 in 2000. On the other hand, the coefficients associated with distance to the nearest smaller cities and places exhibit much larger structural change across time for the larger neighborhood extent. This is particularly true for distance to the nearest place of 10,000 population, which did not significantly influence patterns of sprawl in 1973; however, in 2000 we find that a one kilometer increase in distance increases the sprawl measure by 0.85.

Lastly, Table 11 reports the results of the third model estimated separately for large urban, suburban and exurban counties for 2000 using the 5 km^2 neighborhood extent. The results for the large urban model are quite similar to the full model reported in Table 9. The overall model fit is 0.39 (compared to 0.38 for the full model) and with the exception of the flood indicator variable, all the coefficients are of the same sign and significance level. By comparison, the overall fit for the suburban model is higher (0.51) and for the exurban model is much lower (0.07). Some striking differences in parameter estimates are apparent as well. In addition to several differences in the soil and typography coefficients, the influence of distance to Washington D.C. flips from positive to negative in both the suburban and exurban models. This suggests that instead of being a monotonic function of distance from Washington D.C., residential scatteredness decreases significantly with distance from Washington D.C. In contrast and for both the suburban and exurban models, distance from Baltimore remains positive and highly significant with a coefficient that is much larger in absolute magnitude.

Conclusions

This paper has attempted to shed light on two related research issues that are paramount to measuring and modeling urban sprawl: first, an examination of data and scale dependencies is conducted by comparing measures of spatial pattern across different datasets and across spatial scales of analysis; and second, the ability of the monocentric model to explain spatial variations in residential land use patterns across an urban-rural gradient is tested.

We find convincing evidence of substantial data and scale dependencies. Our investigation of the National Land Cover Dataset (2001) finds that these data systematically misclassify the vast majority of isolated houses that make up low density

[12] Given the linear-log specification of the model, the marginal effect of a one kilometer change in distance on the average measure of residential sprawl is calculated as $\partial y / \partial d = \hat{\beta} / \bar{d}$ where y is the dependent variable (measure of sprawl), d is the distance variable, $\hat{\beta}$ is the estimated coefficient associated with distance and $\bar{d}$ is the average value of the sample.

sprawl. In comparing the spatial location of residential land use as recorded by data from the Maryland Department of Planning (MDP) in 2000, we find that a full 79 percent of the land designated as low density residential land use by the MDP data are coded as undeveloped by the NLCD data. The discrepancy goes beyond definitional differences in what constitutes developed land and whether low density development corresponds to the land parcel boundary or just the built portions. The remotely sensed data miss substantial portions of low density development, including the built structures, and appear to be seriously biased towards misclassifying development that is not surrounded closely by other development. Our analysis using footprint data from Howard County illustrates that the NLCD data miss about 30 percent of all residential structures in the county and demonstrates that the NLCD data systematically fail to record low density residential development that is more scattered.

The implication of these results is that such data cannot be used to reliably record the spatial pattern of residential land use and much less, to measure the incidence of sprawl. Our analysis of pattern using the MDP and NLCD data is consistent with this intuition. We find striking differences in the scatteredness of residential pattern as represented by these two datasets that are not fully explained by their difference in spatial resolution and difference in how developed land is identified and classified. Even though the MDP data are subject to a lower spatial resolution, we find that they convey more of the scattered nature of residential patterns in our study area than do the NLCD data. This is striking, given the fine-scale resolution of the NLCD data and provides further evidence of these data limitations in measuring residential patterns. Taken together, these results cast doubt on the validity of using such data to draw conclusions regarding residential scatteredness or sprawl.

Our pattern analysis reveals important qualitative differences in pattern across the urban-rural continuum that are masked by aggregation of the pattern measure to the Maryland region. We find very noticeable shifts over time, for example in the pattern of residential land use in both suburban and exurban counties that are masked at the state level due to the overweighting of residential pattern in higher density areas. In contrast to the pattern in large urban areas, which remained constant over the time period we study, changes in the pattern distribution in suburban and exurban areas suggest substantial increases in moderate residential scatteredness.

These differences are borne out by subsequent regression analyses that reveal important differences in the monocentric model's ability to explain the observed variation in residential land use pattern across an urban-rural continuum. The model performs remarkably well when applied to residential patterns in urban and in particular suburban areas, explaining over half of the observed variation in pattern in suburban areas while controlling for key sources of local spatial heterogeneity. However, it is exceedingly poor at explaining observed variations in pattern in exurban areas. In addition, we find significant structural change in parameter values across the different scales of analysis. Our results suggest that the monocentric model provides a better explanation of residential sprawl patterns at a more aggregate scale of analysis than a local one and that even at a more aggregate scale, a substantial amount of the variation in pattern is not explained by this model.

Taken together, an implication of our findings is that a different dynamic may be at work in determining the fine-scale spatial pattern of residential land use in exurban areas. Interestingly, the monocentric model appears to have some sway in explaining the fine-scale pattern of residential land use in urban and particularly suburban areas, but its relevance to explaining fine-scale residential pattern in exurban areas appears to be much more limited. This is not to imply that the general notion of urban proximity isn't important to understanding growth in exurban areas; clearly it is, but our findings suggest that this model does not shed light on the *fine-scale pattern* of residential land use in these areas. Thus we are challenged to develop alternative hypotheses to explain the evolution of pattern in these areas. Given the inability of remotely sensed data to identify this type of development and the general dearth of fine-scale land use data available for anything greater than a county or two, this challenge is further complicated by substantial data limitations.

References

Alperovich, G. 1995. The effectiveness of spline urban density functions: An empirical investigation. *Urban Studies* 32: 1537–1548.

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E. 1976. A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964. Available online: http://landcover.usgs.gov/pdf/anderson.pdf

Auch, R., Taylor, J. and Acevedo, W. 2004. Urban Growth in American Cities: Glimpses of U.S. Urbanization. U.S. Geological Survey Circular 1252, January 2004. Available online: http://pubs.usgs.gov/circ/2004/circ1252/.

Brueckner, J.K. 1986. A switching regression analysis of urban population densities. *Journal of Urban Economics* 19: 174–189.

Burchfield, M., Overman, H., Puga, D. and Turner, M. 2006. Causes of Sprawl: A Portrait from Space." *Quarterly Journal of Economics*. May 2006: 587-633.

Carrion-Flores, Carmen and Elena G. Irwin, 2004. "Determinants of Residential Land Use Conversion and Sprawl at the Rural-Urban Fringe." *American Journal of Agricultural Economics*, 86(4): 889-904.

Cutsinger, J. Galster, G., Wolman, H., Hanson, R., and Towns, D., 2005, Verifying the Multi-dimensional Nature of Metropolitan Land Use: Advancing the Understanding and Measurement of Sprawl. *Journal of Urban Affairs*, 27(3), 235-259.

Robert J. Gilliom *and* Gail P. Thelin, Classification and Mapping of Agricultural Land for National Water-Quality Assessment, U.S. Geological Survey Circular 1131; 1997. Available online: http://ca.water.usgs.gov/pnsp/circ1131/data.html

Homer, C. C. Huang, L. Yang, B. Wylie and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. *Photogrammetric Engineering and Remote Sensing* Vol. 70, No. 7, July 2004, pp. 829-840.

Jordon, S., Ross, J. and Usowski, K. 1998. U.S. Suburbanization in the 1980s. *Regional Science and Urban Economics* 28: 611-627.

Luck, M. and Wu, J., 2002, A Gradient Analysis of Urban Landscape Pattern: A Case Study from the Phoenix Metropolitan Region, Arizona, USA. *Landscape Ecology*, 17, 327-339.

McCauley, M. and S. Goetz 2003. Mapping Residential Density Patterns Using Multi-Temporal Landsat Data and a Decision-Tree Classifier, *Int. J. Remote Sens*. 24 (00), 1–19.

McMillen, D. 1994. Vintage growth and population density: an empirical investigation. *Journal of Urban Economics* 36: 333-352.

Mills, E.S.1970. Urban density functions. *Urban Studies* 7:5-20.

Seto, K.C. and Fragkias, M., 2005. Quantifying Spatiotemporal Patterns of Urban Land-Use Change in Four Cities of China with Time Series Landscape Metrics. *Landscape Ecology*, 20, 871-888.

Tsai, Y-H., 2005. Quantifying Urban Form: Compactness versus 'Sprawl.' *Urban Studies*, 42(1), 141-161.

Turner, M. G. 1989. Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics* 20:171-197.

Appendix A: Description of Data

The following provides a general description of the datasets used in the analyses. The basic information listed is from each of the metadata files that are associated with each of the datasets. The GIRAS data description is augmented by the discussions found in the Data Users Guide 4 (http://edc.usgs.gov/geodata/LULC/LULCDataUsersGuide.pdf) and in Anderson, et al. (1976). In addition, we list any known data limitations or considerations and the source for this additional information.

1. 1973 Land Use/Land Cover for Maryland (MDP 1973)

Data generation

Originating Agency: Maryland Department of Planning

Projection and Units: Maryland State Plane Coordinates - meters

Datum: NAD 1983

Minimum mapping unit: 10 acres

Developed using uncontrolled high altitude aerial photography and State Highway Administration 1"=1 mile County Maps.

Original Collection (developed mid 1970s): Various land use categories were identified as polygons and displayed manually as ink on mylar with number codes.

Digitization done in 2003

Additional considerations of 1973 data relative to 2000 Land Use/Land Cover:
(from the metadata file, MDP 1973)

(1) More interior water bodies (e.g. ponds) mapped in 2000 and not in 1973 resulting in an increase in water for 2002

(2) National Wetland Inventory Data used in 2000 to confirm and refine the shape of wetland areas 5 acres or greater in size. These were not used in 1973, which resulted in more accurate mapping of wetlands for 2000.

(3) Minimum mapping unit in 1973: 10 acres; in 2000 additional data from MD Property View were used to refine urban land use categories, resulting in more refined mapping results.

(4) Some transportation features mapped in 2000 (particularly in Baltimore City) and not in 1973 resulting in a decrease in other land use categories in 1973

(5) Code 24 was originally used in 1973 and changed to either 241 or 242 for 2000, resulting in a more comparable dataset.

Land use classification

The land use\land cover classification scheme described below was used to identify the predominant usage of land that could be interpreted from high altitude, uncorrected aerial photography circa 1973. The LU_CODE field, in each county land use shape file, contains the 2 or 3 digit integer numbers identified below. Only land uses greater than 10 acres in size were identified when this map dataset was developed in the mid 1970s. Transportation features such as roads, highways, rail lines and utility lines were not included.

10   Urban Built-Up

11   Low-density residential - Detached single-family/duplex dwelling units, yards and associated areas. Areas of more than 90 percent single-family/duplex dwelling units, with lot sizes of less than five acres but at least one-half acre (.2 dwelling units/acre to 2 dwelling units/acre).

12   Medium-density residential - Detached single-family/duplex, attached single-unit row housing, yards, and associated areas. Areas of more than 90 percent single-family/duplex units and attached single-unit row housing, with lot sizes of less than one-half acre but at least one-eighth acre (2 dwelling units/acre to 8 dwelling units/acre).

13   High-density residential - Attached single-unit row housing, garden apartments, high-rise apartments/condominiums, mobile home and trailer parks. Areas of more than 90 percent high-density residential units, with more than 8 dwelling units per acre.

14   Commercial - Retail and wholesale services. Areas used primarily for the sale of products and services, including associated yards and parking areas.

15   Industrial - Manufacturing and industrial parks, including associated warehouses, storage yards, research laboratories, and parking areas.

16   Institutional - Elementary and secondary schools, middle schools, junior and senior high schools, public and private colleges and universities, military installations (built-up areas only, including buildings and storage, training, and similar areas), churches, medical and health facilities, correctional facilities, and government offices and facilities that are clearly separable from the surrounding land cover.

17   Extractive - Surface mining operations, including sand and gravel pits, quarries, coal surface mines, and deep coal mines. Status of activity (active vs. abandoned) is not distinguished.

18   Open urban land - Urban areas whose use does not require structures, or urban areas where non-conforming uses characterized by open land have become isolated. Included are golf courses, parks, recreation areas (except areas associated with schools or other institutions), cemeteries, and entrapped agricultural and undeveloped land within urban areas.

20   Agriculture

21   Cropland - Field crops and forage crops.

22   Pasture - Land used for pasture, both permanent and rotated; grass.

23   Orchards/vineyards/horticulture - Areas of intensively managed commercial bush and tree crops, including areas used for fruit production, vineyards, sod and seed farms, nurseries, and green houses.

24   Feeding operations - Cattle feed lots, holding lots for animals, hog feeding lots, poultry houses, and commercial fishing areas (including oyster beds).

241 Feeding operations - Cattle feed lots, holding lots for animals, hog feeding lots, poultry houses.

242 Agricultural building breeding and training facilities, storage facilities, built-up areas associated with a farmstead, small farm ponds, commercial fishing areas.

25 Row and garden crops - Intensively managed truck and vegetable farms and associated areas.

40 Forest

41 Deciduous forest - Forested areas in which the trees characteristically lose their leaves at the end of the growing season. Included are such species as oak, hickory, aspen, sycamore, birch, yellow poplar, elm, maple, and cypress.

42 Evergreen forest - Forested areas in which the trees are characterized by persistent foliage throughout the year. Included are such species as white pine, pond pine, hemlock, southern white cedar, and red pine.

43 Mixed forest - Forested areas in which neither deciduous nor evergreen species dominate, but in which there is a combination of both types.

44 Brush - Areas which do not produce timber or other wood products but may have cut-over timber stands, abandoned agriculture fields, or pasture. These areas are characterized by vegetation types such as sumac, vines, rose, brambles, and tree seedlings.

50 Water - Rivers, waterways, reservoirs, ponds, bays, estuaries, and ocean.

60 Wetlands - Forested or non-forested wetlands, including tidal flats, tidal and non-tidal marshes, and upland swamps and wet areas.

70 Barren land

71 Beaches - Extensive shoreline areas of sand and gravel accumulation, with no vegetative cover or other land use.

72 Bare exposed rock - Areas of bedrock exposure, scarps, and other natural accumulations of rock without vegetative cover.

73 Bare ground - Areas of exposed ground caused naturally, by construction, or by other cultural processes.

2. 2000 Land Use/Land Cover for Maryland (MDP 2000)

Data generation

Originating Agency: Maryland Department of Planning

Projection and Units: Maryland State Plane Coordinates - meters

Datum: NAD 1983

Initially developed using high altitude aerial photography and satellite imagery. Urban land use categories were further refined using parcel data from MDPropertyView.

Minimum Mapping Unit: 10 acres

Land use classification

The land use\land cover classification scheme described below has been used to identify the predominant usage of land that could be interpreted from high altitude aerial photography and satellite imagery. The LU_CODE field, in each county land use shape file, contains the 2 or 3 digit integer numbers identified below. In general, only land uses greater than 10 acres in size have been identified, although urban land is identified at a smaller scale by comparisons with parcel data from MDPropertyView. Transportation features such as roads, highways, rail lines and utility lines have not been included in this GIS database. Transportation features are better represented by the point and line files available from the Maryland State Highway Administration.

10 Urban Built-Up

11 Low-density residential - Detached single-family/duplex dwelling units, yards and associated areas. Areas of more than 90 percent single-family/duplex dwelling units, with lot sizes of less than five acres but at least one-half acre (.2 dwelling units/acre to 2 dwelling units/acre).

12 Medium-density residential - Detached single-family/duplex, attached single-unit row housing, yards, and associated areas. Areas of more than 90 percent single-family/duplex units and attached single-unit row housing, with lot sizes of less than one-half acre but at least one-eighth acre (2 dwelling units/acre to 8 dwelling units/acre).

13 High-density residential - Attached single-unit row housing, garden apartments, high-rise apartments/condominiums, mobile home and trailer parks. Areas of more than 90 percent high-density residential units, with more than 8 dwelling units per acre.

14 Commercial - Retail and wholesale services. Areas used primarily for the sale of products and services, including associated yards and parking areas.

15 Industrial - Manufacturing and industrial parks, including associated warehouses, storage yards, research laboratories, and parking areas.

16 Institutional - Elementary and secondary schools, middle schools, junior and senior high schools, public and private colleges and universities, military installations (built-up areas only, including buildings and storage, training, and similar areas), churches, medical and health facilities, correctional facilities, and government offices and facilities that are clearly separable from the surrounding land cover.

17 Extractive - Surface mining operations, including sand and gravel pits, quarries, coal surface mines, and deep coal mines. Status of activity (active vs. abandoned) is not distinguished.

18 Open urban land - Urban areas whose use does not require structures, or urban areas where non-conforming uses characterized by open land have become isolated. Included are golf courses, parks, recreation areas (except areas associated with schools or other institutions), cemeteries, and entrapped agricultural and undeveloped land within urban areas.

20 Agriculture

21 Cropland - Field crops and forage crops.

22 Pasture - Land used for pasture, both permanent and rotated; grass.

23 Orchards/vineyards/horticulture - Areas of intensively managed commercial bush and tree crops, including areas used for fruit production, vineyards, sod and seed farms, nurseries, and green houses.

24 Feeding operations - Cattle feed lots, holding lots for animals, hog feeding lots, poultry houses, and commercial fishing areas (including oyster beds).

241 Feeding operations - Cattle feed lots, holding lots for animals, hog feeding lots, poultry houses.

242 Agricultural building breeding and training facilities, storage facilities, built-up areas associated with a farmstead, small farm ponds, commercial fishing areas.

25 Row and garden crops - Intensively managed truck and vegetable farms and associated areas.

40 Forest

41 Deciduous forest - Forested areas in which the trees characteristically lose their leaves at the end of the growing season. Included are such species as oak, hickory, aspen, sycamore, birch, yellow poplar, elm, maple, and cypress.

42 Evergreen forest - Forested areas in which the trees are characterized by persistent foliage throughout the year. Included are such species as white pine, pond pine, hemlock, southern white cedar, and red pine.

43 Mixed forest - Forested areas in which neither deciduous nor evergreen species dominate, but in which there is a combination of both types.

44 Brush - Areas which do not produce timber or other wood products but may have cut-over timber stands, abandoned agriculture fields, or pasture. These areas are characterized by vegetation types such as sumac, vines, rose, brambles, and tree seedlings.

50 Water - Rivers, waterways, reservoirs, ponds, bays, estuaries, and ocean.

60 Wetlands - Forested or non-forested wetlands, including tidal flats, tidal and non-tidal marshes, and upland swamps and wet areas.

70 Barren land

71 Beaches - Extensive shoreline areas of sand and gravel accumulation, with no vegetative cover or other land use.

72 Bare exposed rock - Areas of bedrock exposure, scarps, and other natural accumulations of rock without vegetative cover.

73 Bare ground - Areas of exposed ground caused naturally, by construction, or by other cultural processes.

80 Transportation - Miscellaneous Transportation features not elsewhere classified.

3. National Land Cover Dataset 2001 (NLCD 2001)

Data generation

Originating Agency: U.S Geological Survey (USGS)

Projection and Units: Albers Conical Equal Area – meters

Datum: NAD 1983

Compiled from Landsat satellite TM imagery with a spatial resolution of 30 meters and supplemented by various ancillary data (where available).

Minimum Mapping Unit: 30 meters x 30 meters

Land use classification

11 Open Water - All areas of open water, generally with less than 25% cover of vegetation or soil.

12 Perennial Ice/Snow - All areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.

21 Developed, Open Space - Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes

22 Developed, Low Intensity - Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units.

23 Developed, Medium Intensity - Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units.

24 Developed, High Intensity - Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80 to100 percent of the total cover.

31 Barren Land (Rock/Sand/Clay) - Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.

32 Unconsolidated Shore* - Unconsolidated material such as silt, sand, or gravel that is subject to inundation and redistribution due to the action of water. Characterized by substrates lacking vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable. Erosion and deposition by waves and currents produce a number of landforms representing this class.

41 Deciduous Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.

42 Evergreen Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.

43 Mixed Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.

51 Dwarf Scrub - Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.

52 Shrub/Scrub - Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.

71 Grassland/Herbaceous - Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.

72 Sedge/Herbaceous - Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.

73 Lichens - Alaska only areas dominated by fruticose or foliose lichens generally greater than 80% of total vegetation.

74 Moss - Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.

81 Pasture/Hay - Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.

82 Cultivated Crops - Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled.

90 Woody Wetlands - Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

91 Palustrine Forested Wetland* -Includes all tidal and non-tidal wetlands dominated by woody vegetation greater than or equal to 5 meters in height and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent.

92 Palustrine Scrub/Shrub Wetland* - Includes all tidal and non-tidal wetlands dominated by woody vegetation less than 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent. The species present could be true shrubs, young trees and shrubs or trees that are small or stunted due to environmental conditions.

93 Estuarine Forested Wetland* - Includes all tidal wetlands dominated by woody vegetation greater than or equal to 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent. Total vegetation coverage is greater than 20 percent.

94 Estuarine Scrub/Shrub Wetland* - Includes all tidal wetlands dominated by woody vegetation less than 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent. Total vegetation coverage is greater than 20 percent.

95 Emergent Herbaceous Wetlands - Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

96 Palustrine Emergent Wetland (Persistent)* - Includes all tidal and non-tidal wetlands dominated by persistent emergent vascular plants, emergent mosses or lichens, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Plants generally remain standing until the next growing season.

97 Estuarine Emergent Wetland* - Includes all tidal wetlands dominated by erect, rooted, herbaceous hydrophytes (excluding mosses and lichens) and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and that are present for most of the growing season in most years. Perennial plants usually dominate these wetlands.

98 Palustrine Aquatic Bed* - The Palustrine Aquatic Bed class includes tidal and nontidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is below 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, detached floating mats, and rooted vascular plant assemblages.

99 Estuarine Aquatic Bed* - Includes tidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, kelp beds, and rooted vascular plant assemblages.

* Coastal NLCD class only

Data considerations

Classification errors: it is difficult to discriminate low-density residential areas due to the range of land cover types within this specific land use and their associated spatial variability. Remotely sensed data will undercount low density development due to classification errors (McCauley and Goetz (2003).

4. Geographic Information Retrieval and Analysis System Land Use and Land Cover Data (GIRAS 1976)

Data generation

Name of Data Set: Originating Agency: U.S Geological Survey (USGS)

Projection and Units: Universal Transverse Mercator - meters

Compiled primarily from color-infrared aerial photography acquired during the mid-1970's (1976 being the median and modal year).

From this photography, polygons of land use and land cover were manually interpreted and delineated on 1:250,000-scale maps, and in a few cases, 1:100,000-scale maps for the entire United States.

Minimum mapping unit: 10 acres for urban uses, as well as for some other categories such as confined feeding operations and strip mines, and 40 acres for all other classes

Land use classification

1 Urban or built-up land
- 11 Residential
- 12 Commercial and services
- 13 Industrial
- 14 Transportation, communication, utilities
- 15 Industrial and commercial complexes
- 16 Mixed urban or built-up land
- 17 Other urban or built-up land

2 Agricultural land
- 21 Cropland and pasture
- 22 Orchards, groves, vineyards, nurseries, and ornamental horticultural
- 23 Confined feeding operations
- 24 Other agricultural land

3 Rangeland
- 31 Herbaceous rangeland
- 32 Shrub and brush rangeland
- 33 Mixed rangeland

4 Forest land
- 41 Deciduous forest land
- 42 Evergreen forest land
- 43 Mixed forest land

5 Water
- 51 Streams and canals
- 52 Lakes
- 53 Reservoirs
- 54 Bays and estuaries

6 Wetland
- 61 Forested wetland

62 Nonforested wetland

7 Barren land
71 Dry salt flats
72 Beaches
73 Sandy areas not beaches
74 Bare exposed rock
75 Strip mines, quarries, gravel pits
76 Transitional areas

8 Tundra
81 Shrub and brush tundra
82 Herbaceous tundra
83 Bare ground
84 Wet tundra
85 Mixed tundra

9 Perennial snow or ice
91 Perennial snowfields
92 Glaciers

Data considerations

Lack of specificity on land-use characteristics, such as dominant crops grown and irrigation (Gilliom and Thelin, 1997), and varying residential classes, such as low, medium and high residential development.

Problem of scale input - These data were originally digitized at a scale of 1:250,000 with some portions of coverage at 1:100,000. Therefore, use of these boundaries with larger scale data is not recommended as it would be beyond the resolution capabilities of the data set (metadata of Arc-Info version of GIRAS, available at http://www.epa.gov/Region8/gis/data/r8_huc.html).

Appendix B: Urban-rural continuum categories

The following definitions were used to classify each Maryland county along an urban-rural continuum:

Large urban: County is located in large metro area (over 2 million) and county population is 500,000 or greater.

Small urban: (1) County is located in small metro area (100,000-250,000 population) and has a county population of less than 100,000 or (2) county is located in a micropolitan area.

Suburban: County is located in large metro area (over 2 million) and county population is less than 500,000 and greater than 150,000.

Exurban: (1) County is located in large metro area (over 2 million) and county population is less than 150,000; or (2) county is located in a small metro area (100-250,000 population) and county population is 100,000-150,000.

Rural: all other counties.