2010 Urban Area Criteria Research at the Census Bureau

Kevin Hawley

ABSTRACT: The Census Bureau has been researching several potential changes to the urban area delineation criteria. Urban areas are delineated in an automated environment each decade. Potential considerations include the use of land use datasets and digital elevation models to identify territory, such as wetlands and areas of steep slopes, that offset urbanization and are taken into consideration during urban area delineation, as well as to identify non-residential urban land use that might not otherwise be included in the urban area because of the criteria's emphasis on residential population density. In addition, the urban area delineation code is being converted from Avenue to Visual Basic using ArcObjects. This presentation will focus primarily on issues relating to the spatial data and spatial analysis tools necessary to delineate urban areas. Some consideration also will be given to aspects of the urban landscape that cannot be easily accounted for based on current spatial datasets or technical limitations.

KEYWORDS: Urban areas, land cover, impervious, steep slopes, wetlands

Introduction

The Census Bureau has used the terms "urban" and "rural" in data products since the 1880 census. The definition of urban has changed throughout the years, progressing from the identification of incorporated places of 2,500 or more population as urban, through adoption of densely settled urbanized areas (UAs) of 50,000 or more people for the 1950 census, in addition to urban places located outside UAs, to adoption of urban clusters (UCs) of at least 2,500 and less than 50,000 people for Census 2000. Both UAs and UCs are defined based largely on residential population density measured at the census block and block group levels. Both types of areas are referred to generically as "urban areas." Territory not designated as urban is defined as rural. The urban areas delineated for the 2000 Census were the first to be created using a nearly fully automated process.

The 2000 urban area delineation built initial cores using population density at the census block group and census block level. Initially, the cores included areas with population densities greater than 1,000 people per square mile (ppsm). Additional block groups and blocks were then iteratively added as contiguous areas with population densities greater than 500 ppsm. Additional processing included filling in enclaves and indentations and connecting nearby cores via hops or jumps using the road network contained within the Topologically Integrated Geographic Encoding and Referencing system (TIGER) (*Federal Register*, 2002).

The potential changes to urban area delineation criteria that will be discussed in this paper make use of ancillary data such as landuse data and Digital Elevation Models (DEM). These potential changes affect the initial core delineation and the hop and jump phases of the urban area delineation.

Using Impervious Surface in the Initial Core Delineation of Urban Areas

In the 2000 urban area delineation program, initial cores were defined based solely on population density, area and contiguity. The first step was to identify all block groups with a population density of 1,000 ppsm or more and an area less than 2 square miles. Any contiguous block groups with a population density of 500 ppsm or more and an area less than 2 square miles was also included. This process of adding contiguous block groups is repeated until no additional block groups qualify. At this point, the initial delineation moves from the block group to the higher resolution block. The block is the highest level of resolution for which the Census Bureau publishes data. Any blocks with a population density greater than 500 ppsm that are contiguous to the block groups that have qualified in previous steps are now added to the core in an iterative fashion. When all contiguous blocks with population density greater than 500 ppsm have been added, the process starts from the beginning using blocks rather than block groups (*Federal Register*, 2002).

The process used to build the initial cores is highly dependent on residential population. Many of the commercial and industrial urban landuse areas with low residential population densities will be captured in the intital core building phase when tracts are qualified and included in the urban area; interior commercial urban landuse will be captured as enclaves. Nevertheless, including ancillary information such as land cover may provide an improvement over the 2000 urban area criteria. In 2000, there were recommendations to include commercial or industrial areas regardless of population density, however, due to the lack of a comprehensive and consistent national database that identified these areas this was not possible (*Federal register*, 2002). The Census Bureau is exploring the use of both commuting data and impervious land cover to aid in identifying non-residential urban land use areas that are components of the urban landscape, but do not have high residential population densities.

Place-of-work data at the census tract level can be used in the initial delineation by aiding in the identification of employment zones that are part of the urban landscape. Place-of-work data can be used to identify the approximate number of workers employed in a particular census tract. For purposes of our research, we have used the tract-to-tract worker flow files from Census 2000. By summing all residential flows to a particular tract, one can calculate the total number of workers in that tract. This allows for the calculation of worker densities at the tract level. In order to make use of these data, the delineation of initial cores would need to reference census tracts rather than block groups, as was the case with the 2000 criteria. The population density criteria would remain the same, but worker density would also be referenced. This will fill in many areas that would otherwise need to be processed during the enclave portion of the delineation. The use of place-of-work data also could help identify employment centers located on the fringes of urban areas that in the past were not accounted for by criteria pertaining to enclaves and indentations.

The use of place-of-work data would represent an improvement over previous decades' definitions, but may still be inadequate in capturing commercial urban land cover on the fringe of an urban area. Because census tracts have specific minimum population requirements, tracts located on the fringes of urban areas generally are larger in land area than those located in more densely populated areas. An industrial area on the fringe of an urban area may have its worker density reduced below the qualifying threshold at the tract level due to the inclusion of area with very few workers. This requires place-of-work data at a higher resolution, but the highest resolution of commuting data is at the tract level. Land cover data, and more specifically impervious land cover, can provide the information necessary to allow the urban area delineation to capture commercial urban land cover on the fringe of urban areas.

The Multi-Resolution Land Characteristics (MRLC) Consortium provides a nationwide dataset of impervious land cover. The MRLC is a group of nine federal agencies that acquired and created land cover datasets collectively known as the National Land Cover Database (NLCD) for the entire United States. Among these land cover datasets is an impervious land cover dataset. Each pixel in the impervious land cover dataset contains a value between 1 and 100 that represents the imperviousness of the pixel as a percentage. The percentages were modeled using a regression tree algorithm based on the relationships between one meter resolution digital orthophotos and Landsat spectral reflectance values (Homer et. al, 2004).

The impervious land cover dataset provides ancillary information that creates a more accurate representation of urban areas by allowing for the inclusion of commercial and industrial urban land cover on the urban fringe. Our research currently uses all pixels in the impervious land cover dataset that have values of 20 or greater. The threshold of twenty percent was chosen because it is a category break between the developed, open space and the developed, low intensity categories as defined in the NLCD 2001 land cover class definitions (Homer et. al, 2004). Setting the threshold on a category break allows one to extract the impervious information from the land cover dataset if necessary. Areas with impervious percentages greater than twenty percent correspond to the following three categories in the land cover dataset: developed, low intensity, developed, medium intensity and developed, high intensity. After the extraction of pixels with imperviousness greater than twenty percent, all of the remaining pixels are classified into a single category and the impervious layer is overlaid with the block layer. The goal is to determine an impervious area percentage for each block as follows:

$$I = 100\left(\frac{A_{i \cap b}}{A_b}\right)$$

where I is the impervious area percentage of the block, $A_{i \cap b}$ is the area of intersection between the impervious land cover and the block, and A_b is the area of the block. Research is currently underway to determine the pros and cons of various methods to arrive at a resulting impervious area percentage. A more time consuming method applies a raster to vector transformation to the impervious layer and performs a union of the new impervious polygons with the blocks. One is then able to calculate the percentage of intersection between the impervious areas and the blocks. Although this method requires longer computing times, the percentages are more reliable. It is also possible to produce results using a generalized impervious polygon layer. Another approach is to use zonal statistics. The blocks are the zones and the zonal statistics produce the area sum of pixels within each zone. This method is generally faster than the vector method, but it requires an internal vector to raster transformation and assigns each cell in the impervious layer to a single cell in the rasterized zone layer. Therefore, a cell on a block boundary will have all of its area transferred to one of the two blocks. The area is not split proportionally between the two blocks. This will result in less reliable percentages and an occasional impervious area percentage greater than 100. After the area percentage for each block has been calculated, blocks with an impervious area percentage greater than 33.3% are flagged as eligible to be included in the urban core as impervious qualifying blocks. Due to the impervious nature of roads, some “median blocks” may get classified as potential impervious qualifying blocks. These blocks are generally very small in area and have large perimeters as they are narrow and are generally bounded on both sides by a separated highway. In order to remove these blocks from inclusion as impervious qualifying blocks, a shape index equation will be applied. The equation is as follows:

$$S = \sqrt{\frac{4\pi A}{P^2}}$$

where S is the shape index, A is the area of the block and P is the perimeter of the block. Blocks with a shape index below a specific threshold will be removed from consideration. Potential impervious qualifying blocks are then added to the initial cores in an iterative fashion based on the contiguity with the initial core. The process continues until there are no potential impervious qualifying blocks that exist that are contiguous with the current initial core.

Figure 1 shows an example of how the use of impervious land cover can improve the boundary of an urban area. The area shown in the figure is Atlantic, Iowa Urban Cluster. The 2000 urban area boundary shows that much of the non-residential urban land use was not included in the urban cluster. The areas to the north did not qualify based on population density, and because there was no way of knowing where the worker densities or commercial and industrial urban land uses were located, this area was excluded. This example also shows that the area is on the urban fringe. Had this area been interior to the urban area boundary, it would have been added as an enclave. Figure 2 shows the impervious land cover along with the new boundary. Using impervious land cover, the

Figure 1: The 2000 UrbanArea for the northern portion of Atlantic, IA overlaid with an aerial photo.

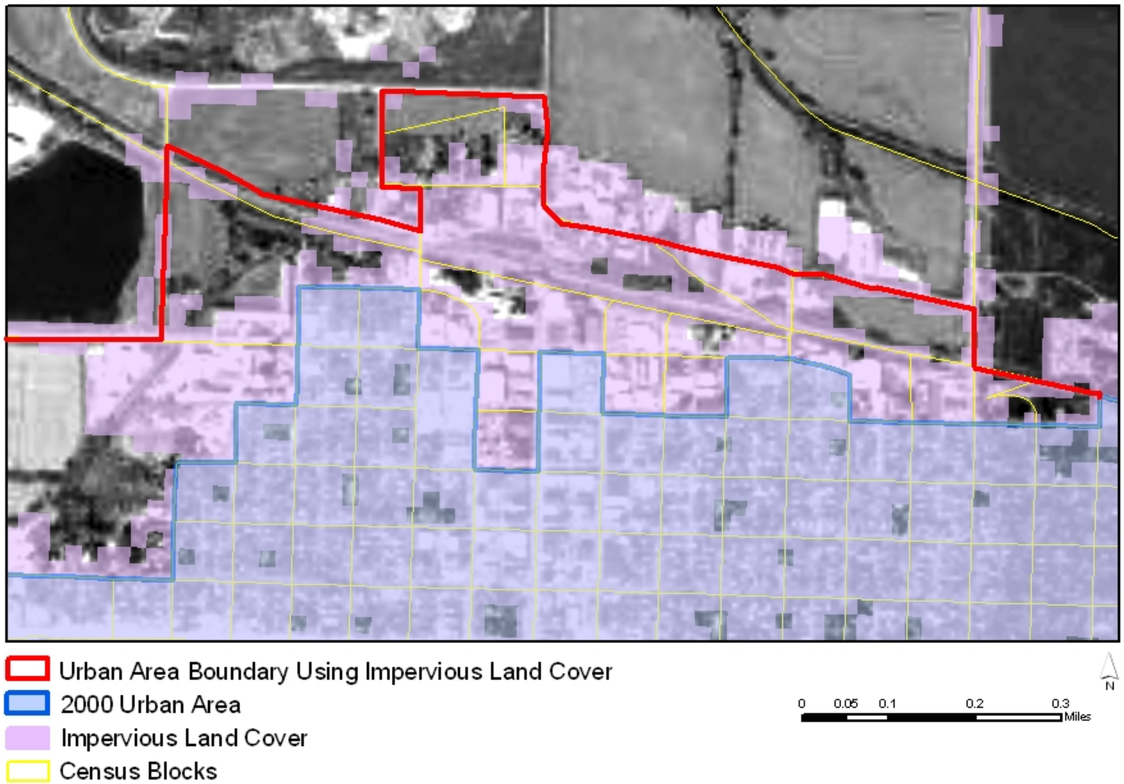

Figure 2: The use of impervious land cover for defining urban areas in Atlantic, IA.

urban area delineation is able to capture much of the urban land cover that was not part of the 2000 urban area. There is a substantial area of impervious land cover just north of the red line in Figure 2 that would not be included in the urban area. This is due to the fact that the block is so large that it only has an impervious area percentage of 12%. This is far below the current threshold of 33.3%.

In addition to being a useful data source for the initial delineation of urban areas, the impervious land cover data may also prove to be useful for intra-Census updates to the boundaries of urban areas. Currently, urban area boundaries are only defined after each decennial census. Updated land cover data could allow for the identification of growth on the urban fringe and updating urban areas in between each census.

Adding Slope and Wetlands as Exempted Territory

In 2000 the Census Bureau considered "water bodies, uninhabited census blocks adjacent to bridged water bodies, military installations, national parks and national monuments" (*Federal Register*, 2002) to be exempted territory (referred to in previous decades' criteria as "undevelopable territory"). Exempted territory becomes important during the hop and jump phases of urban area delineation. Once the initial cores have been built, each core attempts to hop or jump to nearby cores using road connections. Any core with a total population greater than 1,000 may attempt to connect with to nearby cores. The maximum non-exempt distance for a hop is 0.5 miles. However, if the road connection passes through exempted territory, the total distance of the connection may be up to 2.5 miles. After all hops are complete, cores attempt to jump to one another in a similar fashion, but the non-exempt distance for a jump is from 0.5 to 2.5 miles. Any core with a total population greater than 1,500 may attempt to jump to nearby cores. A jump can be up to 5 miles if the road connection passes through exempted territory (Federal Register, 2000).

The reason for recognizing exempted territory is to allow nearby non-contiguous cores to be merged together when land cover restricts development between the two cores. Conceptually, the identification of exempted territory and use in delineation of urban areas is meant to account for instances in which such territory offsets the pattern of urban development, but does not restrict the development of functional ties between two cores (note that tract-to-tract commuting data could be used to indicate when such functional relationships exist). The Census Bureau is currently researching the addition of steep slope areas and wetlands to the list of exempted areas for the 2010 urban area delineation. The Census Bureau previously has taken the presence of wetlands into account when identifying exempted territory, although sources of data have not been consistent nationwide. We have not, however, taken steeply sloped areas into consideration when identifying exempted territory, distinguishing such areas as “undeveloped” rather than “undevelopable”. Nevertheless, we are aware of areas in which urban cores that are separated physically by steeply sloped mountain passes do exhibit close social and economic ties as measured by commuting. The argument typically presented to the Census Bureau is that were it not for the presence of the mountainous topography, the two cores would have grown together over time. Although the TIGER database does not contain steep slope areas or wetlands, these data can now be obtained for the entire nation, and the utility of such data in defining urban areas ca be fully assessed. The National Elevation Dataset (NED) is accessible via the National Map and the wetlands data can be accessed as part of the MRLC land cover dataset.

One of the factors associated with landform that can restrict commercial, industrial and residential development is the slope of the land. Areas of very steep slope make development nearly impossible. This is why steep slope areas are being considered as a potential new addition in the areas defined as exempted areas. The NED is a one arc second (resolution of approximately 30 meters) raster layer that has been assembled by the US Geological Survey (USGS) and is available for the entire United States. Data for Alaska is two arc seconds (USGS Fact Sheet, 1999).

In order to make use of the DEM, slope is calculated. Slope is being calculated using Spatial Analyst functions available in ArcObjects using ArcGIS. The slope is calculated for each cell as the maximum rate of change between the central cell and its neighbors within a 3 x 3 spatial filter using the average maximum technique (ESRI, 2007). The slope is represented as a percentage. After slopes have been calculated, all cells with slope greater than 50 percent are extracted to a new raster layer. The extracted cells are then converted to polygons in a raster to vector transformation. In addition to the polygons with greater than 50 percent slope, small blocks (blocks with an area less than 0.1 square miles and a total population of less than 10) that are within 0.01 miles of steep slope polygons also become exempted areas. This is accomplished by creating a buffer around the small blocks and flagging any blocks that intersect the 50 percent slope areas. The buffer is created around the blocks rather than the slope polygons to save processing time. The amount of time required to buffer the very detailed slope polygons made it unreasonable to buffer the steep slope polygons. After the small blocks within 0.01 miles of the steep slope polygons are identified, the resulting blocks are unioned with the steep slope polygons to create a steep slope exempted territory layer. Figure 3 shows the steep slope areas along with the small area blocks that represent the steep slope exempted territory.

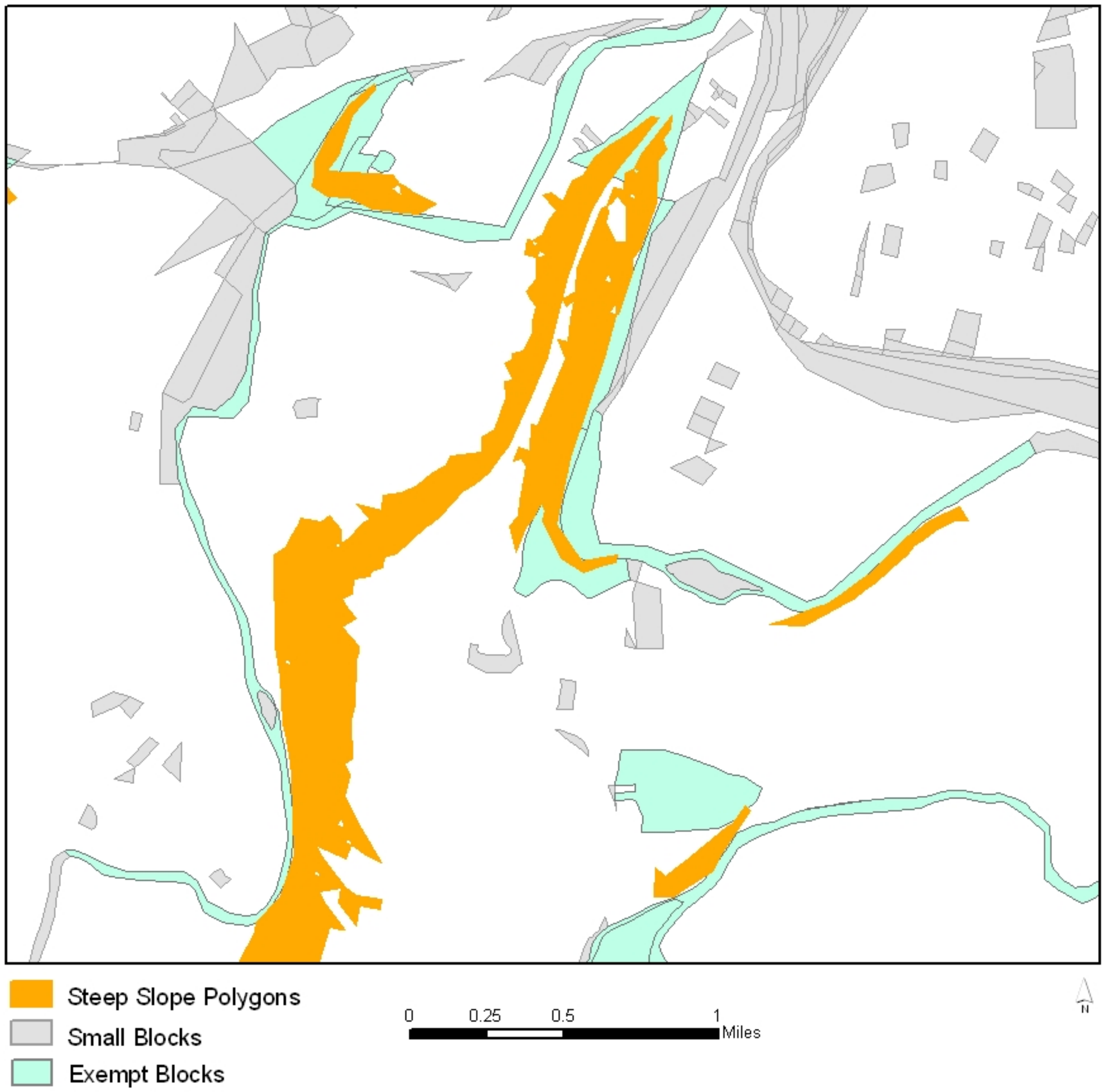

Figure 3: Steep slope areas and small blocks selected as exempted territory.

Another type of land cover that impedes commercial, industrial and residential development is wetlands. Wetlands include marshes, swamps, bogs and fens and greatly contribute to ecological diversity and functional importance such as flood control and pollutant absorbance. Wetlands may or may not contain water year round. For these reasons, development is constricted in these areas. In the United States, the Clean Water Act requires a 404 permit for any development in wetlands (EPA Fact Sheet, 2004).

Wetland areas can be extracted from the land cover datasets provided by the MRLC. The MRLC defines two major categories of wetlands: woody wetlands and emergent herbaceous wetlands. The first step is to extract the cells within the wetland categories from the land cover dataset. The only locations where the hop or jump algorithm of the UA delineation would be making use of wetlands as exempted territory is where a road crosses a wetland. However, in the land cover dataset, a road will be classified as developed land. This means that when the wetland cells are extracted from the land cover database, there will be gaps where the road is located over the wetland. Therefore, there is a need to "fill" in the wetland coverage in these locations. In order to fill in these areas, the ArcGIS Spatial Analyst functions Expand and Shrink are used.

The expand function expands the outer boundary of a particular category of cells in a raster layer by a defined number of cells. The shrink function shrinks the outer boundary of a particular category of cells in a raster layer by a defined number of cells. First, the expand function is run using a cell expansion size of 2. An expansion size of two cells was chosen after research in several wetland areas. In many cases an expansion on one cell was enough to fill in the highways. In reality, an expansion of one cell means that areas that are separated by two cells will be joined together. Since an expansion size of two cells is being used, areas that are separated by four cells (120 meters) will be joined together. Although this may seem like a large gap, it was found that an expansion size of one cell would not suffice for roads that were oriented in a southwest to northeast direction or northwest to southeast direction. Separated highways also required a higher expansion size. This means the outer boundary of the wetland area grows by two cells (60 meters) in every direction. Areas that needed to be filled in will now be merged together as the raster cells have been joined together. However, all other areas now have a 60 meter buffer as the wetlands have expanded into areas where they should not be. The shrink function will now reduce the outer boundary by two cells and in most cases will be unchanged from the original land cover dataset. Figure 4 shows an example of applying the expand and shrink functions on an area where roads exist over wetlands. After the expand and shrink functions have been applied, a raster to vector transformation takes place to create wetlands polygons.

There is still a question as to whether other water areas will be pulled from the TIGER database or from the land cover database. The advantage of using water polygons from TIGER is that their boundaries would be spatially integrated with the block boundaries whereas the boundaries formed from the water areas in the land cover dataset would not. However, using the water areas from the land cover dataset would allow for expansion and shrinking of the wetlands and water areas so that the wetland and water polygons would be integrated with one another even though they would not be completely integrated with the blocks. Small gaps between the wetland layer and TIGER water areas would most likely exist in places where there should not actually be a gap.

All exempted territory will be merged together into a single polygon file when all preliminary processing is complete. The merging will be done via a union operation to be sure that no polygons overlap one another. Whenever a hop or jump is attempted and the distance of the route is greater than the non-exempt allowable distance, the road connection is intersected with the exempted polygon layer and the distance of intersection is calculated.

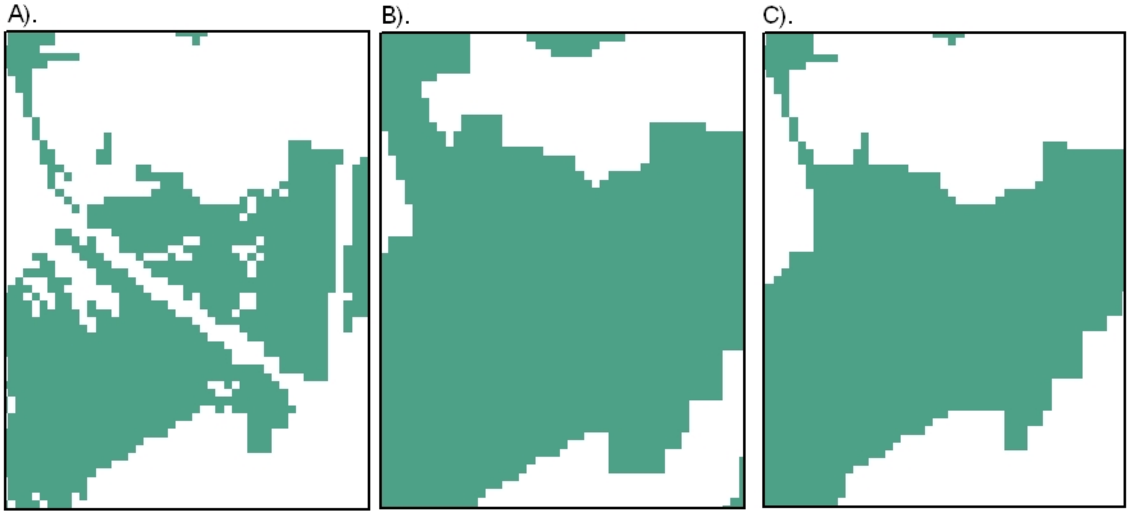

Figure 4: A). Wetland cells extracted from the land cover database. B). Wetlands after application of the expand function with. C). Wetlands after application of the shrink function. This is the result that will be vectorized and contribute to exempted territory.

Moving to ArcGIS

The 2000 urban area delineation software was implemented in three different GIS systems. The initial delineation was implemented in MapInfo's MapBasic. Enclaves and indentations were implemented in ArcInfo using AMLs and hops and jumps were implemented in Avenue in ArcView 3.2. One of the reasons for moving to ArcGIS is that ESRI no longer supports ArcView 3.2. In addition, having all software in a single programming language means one does not need to be fluent in three very different programming languages to update portions of the code.

ArcGIS offers a library of COM objects known as ArcObjects that can be accessed through many different programming languages such as Visual Basic, .NET, C#, C++ and Java. ArcGIS itself is implemented using ArcObjects, so all of the functionality of ArcGIS can be accessed through ArcObjects (Burke et. al, 2003). The UA delineation software for 2010 is being developed in Visual Basic for Applications (VBA) which allows the software to run in an environment where adjustments can easily be made during runtime as the code is not compiled as an executable file. VBA code can be directly altered within ArcGIS, so no additional software needs to be installed. For these reasons, VBA is a particularly attractive choice during the research and development phase of the UA delineation.

Conclusions

Urban areas are defined by the Census Bureau for statistical purposes, to tabulate and present demographic and housing data for urban (and by extension rural) populations. Defining urban areas allows for the presentation of data in an urban-rural dichotomy. Although these areas are defined for statistical purposes, the Census Bureau recognizes that they are used by other governmental entities and organizations to determine eligibility for certain programs. For example, legislation may only offer funding to a hospital that is located in a rural area or a school in an urban area. For both the statistical integrity and program eligibility, it is important that urban areas are defined in an objective way. It is also important to make an attempt to include meaningful data from other sources that can be applied consistently and objectively in the delineation of urban areas. For this reason we are researching methods to improve the delineation process. The inclusion of land cover data makes it possible to access information about the urban landscape (such as the presence or absence of built up areas) that cannot be obtained through the use of block level demographic data such as the presence or absence of built up areas. In addition, land cover data can provide the locations of wetlands that restrict urban development and may be useful to designate as exempted territory. The use of DEMs allows for the calculation of slope which is another factor restricting urban development.

Acknowledgements

Much of the potential criteria discussed in this paper was researched and proposed by Ryan Short. Mr. Short was the main contributor to the development of the 2000 urban area criteria and automated delineation procedures. Without his research efforts, this paper would not have been possible.

REFERENCES

Burke, R., Arana, A. and Tilton, T (2003). Getting to Know ArcObjects. Environmental Systems Research Institute Inc..

ESRI (2007) Surface (Spatial Analyst) – How Slope Works. ArcGIS Desktop Help, Environmental Systems Research Institute Inc.

Homer, C., Huang, C., Yang, L., Wylie, B. and Coan, M. (2004) Development of a 2001 National Land-Cover Database for the United States. ***Photogrammetric Engineering & Remote Sensing.*** 70, 7, pp. 829-840.

EPA Fact Sheet 843-F-04-011a (2004) Wetlands Overview. United States Environmental Protection Agency, http://www.epa.gov/owow/wetlands/facts/contents.html.

Federal Register (2002) Urban Area Criteria for Census 2000; Notice. March 15, 2002. Bureau of the Census, Department of Commerce, Vol. 67, 51, pp. 11663-11670.

USGS Fact Sheet 148-99 (1999) National Elevation Dataset Fact Sheet. United States Geological Survey, Department of the Interior. http://erg.usgs.gov/isb/pubs/factsheets/fs14899.html

Kevin Hawley, Geographer, US Census Bureau, Geographic Standards and Criteria Branch., Washington, DC. E-mail:<kevin.j.hawley@census.gov>.