USGS Next Generation Topographic Map - Graphics Research

West Virginia GIS Technical Center

Final Report

December 2009

Prepared by:

Kevin R Kuhn

Evan Fedorko

West Virginia GIS Technical Center
Morgantown, WV 26506
304.293.9462

Contents

i. Purpose

The West Virginia GIS Technical Center (WVGISTC) at West Virginia University, Morgantown, West Virginia, has completed its investigation for "Development of Automated Graphics Features for Next Generation USGS Topographic Series in Support of the National Map." This report documents the findings investigated for the scope of the project.

Since 1879, 1:24,000 scale topographic maps published by the United States Geologic Survey (USGS) have been the gold standard of reference maps for the United States and its territories. Recent history has seen the rapid emergence of new mapping technologies, and in response to those, the USGS has developed "The National Map," an internet based framework application for the sharing and distribution of up-to-date geospatial data. The days of mass printing of 1:24,000 topographic maps are gone. However, the topographic map is far from irrelevant and remains a central focus of the USGS. This focus has shifted to digital maps but the type of information contained on traditional topographic maps remains relevant. This transition to digital mapping not only applies to the information itself, but how that information is cartographically represented. Two areas of USGS cartographic tradition were the focus of this research project; the red urban tint that is used to reflect built-up areas and the green tint used as a forest cover. As these maps were originally generated manually by skilled cartographers, a new methodology to create these in mass, using digital data and automated processes is needed. Identifying suitable data and creating these processes were the main goals of this research.

The report focuses on use of available large-scale data from state sources for incorporation into the graphics production and use of automated feature extraction and statistical generalization techniques to develop the urban area (red tint) and forest cover (green tints) polygons represented on traditional 1:24,000 USGS quadrangles and for the next generation of USGS topographic map series. For urban designations, the project utilized 1:4,800 scale building centroids, building footprint polygons and road centerlines. For forest cover, the project investigated various combinations of forest land cover types associated with the National Land Cover Database (NLCD 2001) and techniques to improve its usability in a cartographic quality application.

This report and the research performed were based on several USGS technical documents. They are referred to throughout this report as NMP, the abbreviation of the National Mapping Program Technical Instructions (NMP) and Section Part. Specifically, NMP Part 6 covers the Built-up Standards used for the Urban Tint generation and Part 9 details covers the Vegetative Surface Cover used for Forest Cover Tint generation.

ii. Sections

This report is divided into three major sections. Each section describes separate areas of research performed for this investigation. Section 1 is titled Urban Tint Generation, which discusses the methods and data used to generate the USGS Built-Up Areas (red tint). There are two subsections that discusses various approaches researched utilizing road data and utilizing structure data. Section 2 of this report discusses research performed to identify the data and process to generate the USGS Forest Cover or Woodlands (green tint). Section 3 discusses additional cartographic techniques explored for structure data in a digital map environment.

Section 1
Urban Tint Generation

1.0.0 Urban Tint Concepts

The WVGISTC explored several methodologies for creating the USGS urban tint areas using digital processes. The goal of this research was to replicate the original criteria that USGS has used to define and collect urban tint areas. The key elements of creating an urban tint can be rendered down from its cartographic origins (NMP Part 6; Box 1), to identify the correlating GIS elements in its creation using digital processes. Those elements are structures (buildings), their associated use areas(parking lots, etc) and road networks. Using these elements, we are able to create conditions that equate to a built-up area, through representations of structures and road networks.

This research focuses on two distinct aspects of generating an urban tint layer; the first using road centerlines and the second utilizing structure points. The concepts and procedures in this report provide for a repeatable methodology useable by other entities. We discuss the positive and negative outcomes of these processes.

> *BUILT-UP AREA - An area of intensive use, with much of the land covered by structures.*
>
> *The limits of BUILT-UP AREA are determined by the relative concentration of buildings and associated intensive use areas, such as parking lots, and the existence of a systematic street pattern. Density of building and associated use areas will vary from densely concentrated areas downtown, to moderately concentrated residential areas where most of the property is developed.*
>
> *BUILT-UP AREA is > 30% built-up, and is > 40 acres in size, and is > 660 feet along the shortest axis*

Box 1 - Part 6, NMP Technical Instructions

This research used the built-up density criteria and area size criteria to identify candidate areas. A candidate area refers to areas that meet at least one criteria of the capture conditions. For example, the density rasters will identify the first criteria of a 30% built-up density and additional processing will narrow those areas by identifying minimum area size. Further research is needed in order for dimension criteria to be used; this topic is discussed in Section 1.5.5.c

USGS NMP nomenclature uses the terms *Urban Tint* and *Built-Up* Areas interchangeably. This report will distinguish between the two; Built-Up Area refers to a physical condition as described in Box 1. Urban Tint is defined as the desired GIS data layer, or graphic features, which represent Built-Up Areas. Box 1 contains the criteria used to define an Urban Tint and the conditions which it is collected as a layer.

The processes described in this report may refer to automated graphics. Automated graphics is the term used to describe the digital processes and resulting Urban Tint layer researched in this report. During this research these digital processes were performed manually. However, future research could develop programming code or scripting to automate these digital processes.

1.0.1 Data Sources

West Virginia is blessed with several current, highly detailed datasets that will allow for a largely automated derivation of these graphic features. These datasets were created as part of the West Virginia State Addressing and Mapping program's work and include photogrammetrically collected planimetric datasets, including the road and building data used in this research (Figure 1). This data was collected from 2-foot natural color, leaf off photography flown in 2003. Other states may not have this type of data available, but it is expected that a point structure dataset and detailed road dataset will become available as states further develop their GIS spatial data infrastructure. Furthermore, as new programs such as Imagery for the Nation, help produce current, detailed views of the landscape, the methodologies outlined in this investigation will allow urban tint areas to be generated and updated with new information.

Figure 1 - Planimetric Data

1.1.0 Urban Tint from Roads

The first process tested in this research utilized road data to create an urban tint layer. WVGISTC researched the process and tools needed to transform road centerline data into a cartographic polygon (Urban Tint) which could then represent built-up areas. This process established a correlation between road segments and built-up area. Because not all roads may represent a built-up area, the road centerline features were assigned a weighted value based on feature type. This linear feature data was transformed into a raster dataset based on line density and feature weight. The resulting raster was then processed to refine and enhance candidate areas based on NMP criteria. This raster data was then converted to polygon data for further refinement and cartographic enhancement. The process described on the following pages was designed to closely mimic the conditions that USGS uses to collect a Built-up Area and produce an Urban Tint dataset.

1.1.1 Data

In an attempt to provide a repeatable process, we chose to use a nationally available dataset; a subset of the US Census TIGER/Line. The 2003 SAMB 1:4,800-scale road centerline geometry was incorporated into the Census TIGER/Line data as part of a major revision of the TIGER/Line data in 2007-08. Census TIGER/Line data is categorized by multiple use codes (MAF/TIGER Feature class codes; MTFCC). The TIGER/Line data contains numerous features (such as political boundaries) not of use in this study, so the MTFCC codes were used to subset transportation features used in this research. These codes were also used to assign feature weights to the roads (Figure 2). In our study, we used the following feature codes and assigned weights (See Section 1.1.2.b).

MTFCC codes used and weight assigned:

- S1100 (Primary Roads) = 0
- S1730 (Alley) = 1
- S1740 (Private Road) = 1
- S1200 (Secondary Roads) = 2
- S1400 (Local Roads) = 3

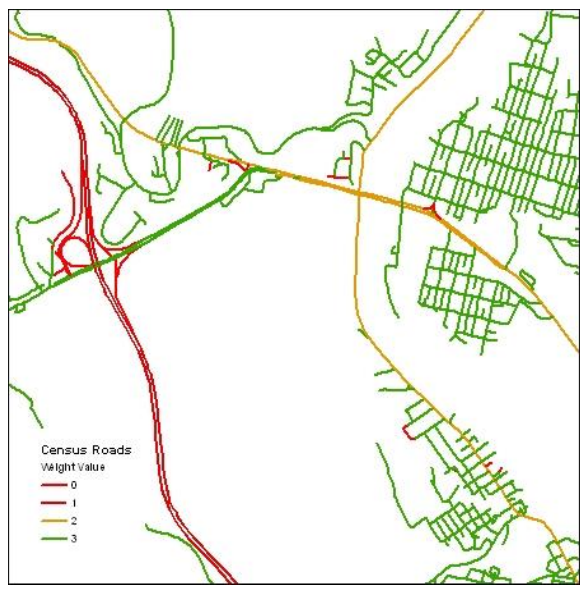

Figure 2 - Census Classification Weight

Though we used Census TIGER/Line data for this research, these concepts are applicable to virtually any road dataset. Testing determined that the type of road plays a role in accurate urban tint generation and care needs to be taken in creating a valid set of input features. Specifically, many transportation datasets contain features such as on-ramps, miscellaneous roads, trails and railroads. These features were excluded from the process as they may not reflect the desired results. Trails do not usually represent built-up areas. Railroads may physically represent developed or built-up areas, but are linear in nature, and therefore not useful in defining specific areas. Rail yards could be applicable, but are specified in the NMP Part6 as areas that are not collected for built-up areas.

1.1.2 Density

We explored road density to assess and identify areas that comply with the NMP Part 6, 30% Built-Up criteria to produce an Urban Tint. The line density tool was used as a direct method of creating a raster surface from the road data. This tool results in a raster dataset with values of density measure: linear unit/per area. Figure 3 shows the source road data on top of its density raster. Low density areas are shown in reds, medium density shown in yellows and high density areas are shown in greens.

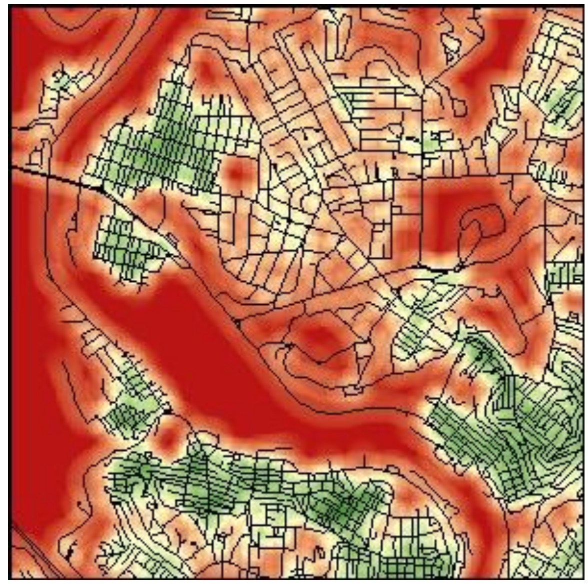

Figure 3 - Density Raster and Roads

Road centerline data is only one dimensional and therefore needs to be assigned an equivalent to an area value. Physical roads in the real world do have an area, measured as the length and width. (Theoretically, detailed centerline road datasets could have this width data; but for this research we standardized road width as part of the weight value (discussed in Section 1.1.2.b). The physical surface area of the road would be considered built-up and would count directly towards the 30% built-up criteria of the area. The built-up criteria also incorporates parking lots and structures as well as roads. Therefore, an additional area value needs to be assumed and assigned to the road segments. Conceptually, this value is not only considering the road length and width, but also the use area surrounding the road on either side.

Testing has shown that the density of the urban core is often close or equal to the density of abutting or adjacent corridors and therefore they are often incorporated into the candidate areas. Conceptually, this makes sense as inner-city roads are not significantly closer together or farther apart than suburban neighborhood roads. This will vary with different areas of development and is also impacted by numerous other variables such as terrain, city code or political boundaries.

There are several factors that influence the outcome of the line density raster, including cell size and feature weight. We discuss the impact of these two variables in the next section.

1.1.2.a Cell size

Two cell sizes were tested: 10 meter and 30 meter. Figure 4 shows the 10 meter cell in grey, the 30 meter cell in tan and the Census TIGER Roads on top. The 10 meter cell size produced a finer resolution raster, as expected, which allowed a finer edge to the urban tint. A potential drawback of using a 10 meter cell size would be longer processing time and heavier hardware requirements for performing the process over large areas. The process will run quicker using the 30 meter cell size, but, the resulting raster has less information (shape and

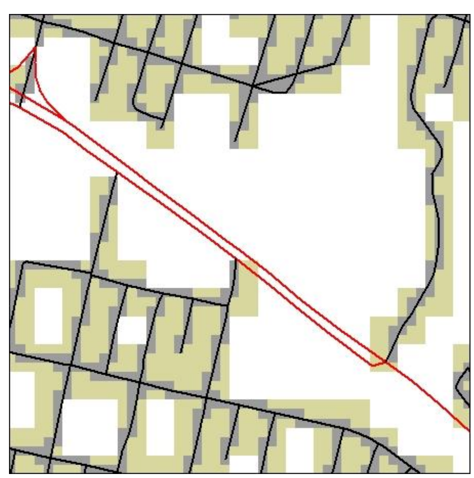

Figure 4 - Cell Size Comparison

boundary detail). This slightly effected the cartographic boundary of the urban tint. Both cell sizes are suitable for creating a density raster, but a preference for 10 meter cell size is given, should computer processing power allow. This research used a 10 meter cell size; subsequent variables and thresholds may need to be adjusted if a 30 meter cell size is used.

1.1.2.b Road Weight by Class

Using Census TIGER road classes, we were able to assign feature weights prior to creating the density raster. A feature weight works by allowing some features to be counted more often than other features. Assigning a weight value based on the roads class allows for a better correlation between a road segment and the surrounding built environment.

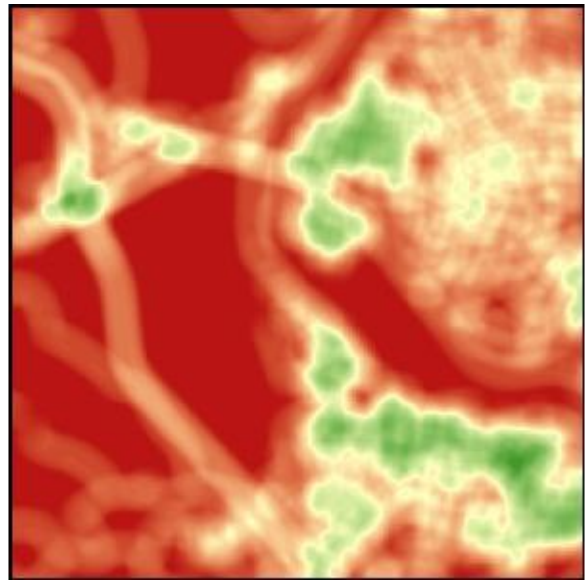

Figure 5 - Road Density Using Non-Weighted Roads

The concept behind the weight scheme is that not all roads equally represent the built-up environment. In general, highways traverse both city and rural areas and do not necessarily represent a built environment. Secondary roads can pass through urban areas, but can also often be rural; therefore they may or may not represent built environments. Local roads are generally found grouped together and often form the urban road network. Local roads are therefore most likely to represent built-up areas.

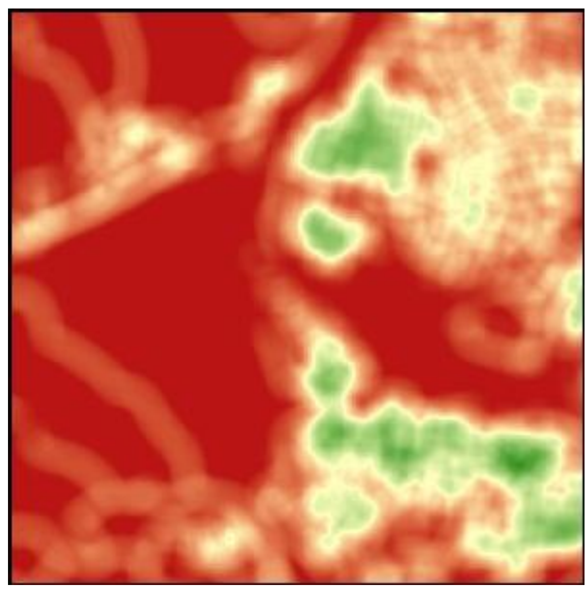

Figure 6 - Road Density Using Weighted Roads

During preliminary tests, density rasters were created from road features without a feature weight. This results with highway features being over-represented as dense areas. Figures 5 & 6 display a density raster generated from road data. Figure 5 clearly shows the inclusion of a highway running from top to bottom on the left of the graphic. Figure 6 shows the same area, but uses road features with assigned weighting. Figure 6 shows similar areas of density, but excludes the highway. The high density areas that result from highways was due because most highways are represented by dual line geometry, which will have a higher density than single line roads. High density which results from highways may often be a false representation. Typical highways would not necessarily reflect a high density area, especially for rural states such as West Virginia. Highways in general were not found to be beneficial in forming clusters of density. Even in dense urban areas, local roads can better represent built-up areas than highways. In addition, the linear nature of highways resulted in elongated, narrow, high density candidate areas. For these reasons, highways were ultimately removed from inclusion when creating the density raster.

For this research, we assigned weight values to features based on the Census TIGER/Line MTFCC codes. The following classifications and weights were assigned: miscellaneous roads-1, secondary roads- 2 and local roads-3 (See Figure 2). These weight values could be altered depending on the data, criteria or area of interest. For example, alleys typically only exist in urban areas and could be more heavily weighted than state routes.

1.1.3 Focal Statistics

The density raster is then processed through a focal statistics function. The focal statistics process refines and enhances the candidate areas by maximizing the larger areas and minimizing the smaller areas. The focal statistics function used is a statistical mean; a statistic that averages the density of a given area (the neighborhood). The neighborhood is defined by the number of cells of the raster that is used to calculate the new value. Focal statistics was able to highlight clusters and minimized the linear habit of the road density. This step is done in part to accommodate the NMP Part 6 Specification to smooth and simplify the Urban Tint boundary (Box 2).

> *The outline of BUILT-UP AREA should be relatively smooth. It should not be extended to include a few buildings nor to exclude a few buildings when the smoothness of the edge would be effected.*
>
> *When the perimeter streets are >30% developed, the BUILT-UP AREA should be extended across the street to include these buildings.*

Box 2 - Part 6, NMP Technical Instructions

The size and shape of the neighborhood function effected the shape and composition of the output candidate areas. In general, the larger the neighborhood size, the more smooth (or averaged) the raster becomes. This was found to be desirable as it removed small undulations of the urban area boundary and enhances key areas (such as larger areas of high density). This process was in keeping with the NMP Part 6 (Box 2).This process must balance between a neighborhood size large enough to minimize the smaller areas, yet keeping the neighborhood size small enough to limit the smoothing effect.

Figures 7-11 demonstrate the effect of altering the Focal Statistics neighborhood size and shape. All figures shown were generated from the same density raster and show the corresponding source road data on top. The figures show the reclassified focal rasters using the same classification threshold, for sake of clarity. The classification threshold is the cell value used to classify cells into a candidate area class or non-candidate area class. Section 1.1.4 will discuss this topic further. Figure 7 shows the original density raster (grey tint), which has not been processed through focal statistics. The figure shows that the density raster has several small island areas apart from the larger areas and exhibits a rough and jagged boundary. These traits are reduced from using focal statistics process.

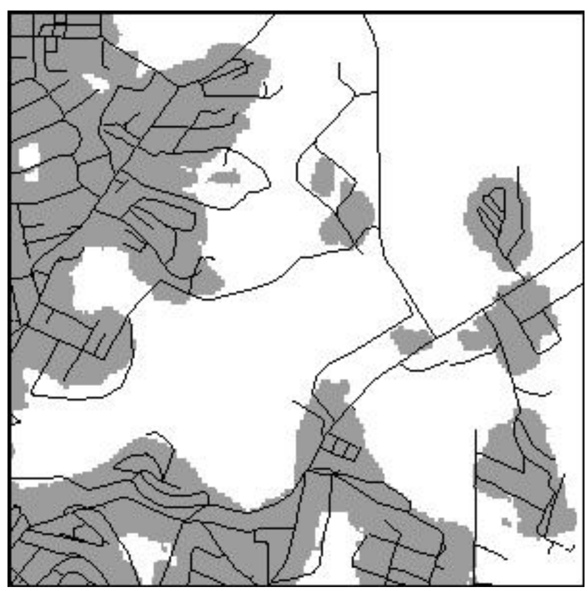

Figure 7 - Original Density Raster

Figure 8 shows the product of a rectangular neighborhood shape. The figure shows that a few of the smaller islands have been removed; this shows the effect of lowering their cell values below the reclassification threshold. Testing found that the annulus (donut) neighborhood shape tended to smooth the boundary areas more than the rectangular shape. Figure 9 demonstrates the focal process using a similar sized neighborhood, but in an annulus shape. This figure shows an even greater smoothing effect and further reduction of small island areas. Neighborhood size also effected the shape and area of the output candidate areas. The larger the neighborhood size, the greater the smoothing effect it has on the output raster. Figure 10 shows a neighborhood radius of 30 cells (300m) than the Figure 9 neighborhood radius of 15 cells (150 meters). Figure 11 shows a smaller neighborhood radius using only a 3 cell radius. This figure demonstrates that few of the island areas have been removed and only slight smoothing has taken place. This research used the annulus neighborhood shape, and a 2 cell inner radius and 15 cell outer radius (Figure 9).

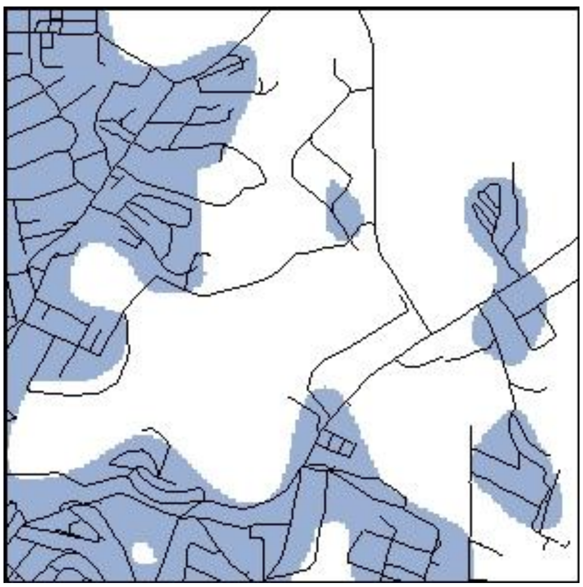

Figure 8 - Focal Raster, Rectangle Neighborhood Shape, 15 cell x 15 cell search size

Figure 9 - Focal Raster, Annulus Neighborhood Shape, 2 cell Inner radius, 15 cell Outer radius

Figure 10 - Focal Raster, Annulus Neighborhood Shape, 2 cell Inner radius, 15 cell Outer radius

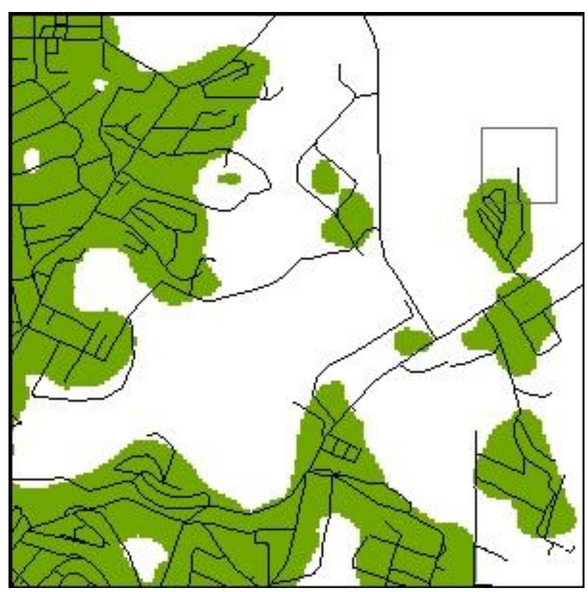

Figure 11 - Focal Raster, Annulus Neighborhood Shape, 2 cell Inner radius, 15 cell Outer radius

1.1.4 Reclassify

Once the focal statistics raster has been calculated, the resulting raster (Figure 12) is then reclassified into two data classes. This reclassification step acts as a threshold to determine which cells are classified as candidate areas and which cells are not classified as candidate areas. The reclassification plays an important role in determining the size of the overall candidate areas. Altering this threshold can expand or shrink the candidate areas. The product of this reclassification step is another raster which is then converted to a polygon dataset.

Our classification threshold value was 1.9. This threshold was chosen because it represented the best balance of inclusion versus exclusion. The resulting classification shown in Figure 13, identifies areas of high density with values of 1.9, shown as yellow. The 1 values, shown in green, and Null values (white) are discarded. After several tests, it was found to be beneficial to be more inclusive during this step. Other processes used later can be used to exclude areas.

Figure 12 - Road Density

Figure 13 - Reclassified Road Density

1.1.5 Cartographic Processes

After the classification step, both the road generated and structure generated tint rasters undergo the same process. Section 1.4.0 describes in further detail the processes used to clean and enhance the data to complete the urban tint layer.

1.1.6 Research Summary

This research demonstrates that existing data and methodologies can be used as a viable means to bridge the past paper topographic maps with the next generation of digital mapping products. The road generated urban tints have stayed true to the original USGS urban tints over the majority of the study area, though there are differences. Most areas of difference can be attributed to areas of genuine new growth, as reflected in the source data, or original tint areas that were interpreted by a technician for inclusion.

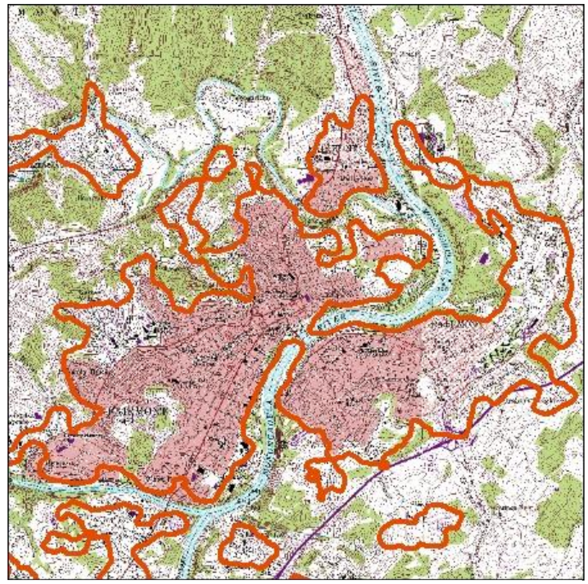

Figure 14 - Generated Tint outline and USGS Topographic map mosaic

Figure 14 shows the generated tint outline in orange on top of a mosaic of USGS topographic maps. This graphic illustrates that the majority of the generated polygons coincide with existing USGS urban tint areas. There are additional areas shown in Figure 14; these were included because they met the same criteria that that is used to define the larger areas. These additional areas could be removed by raising the threshold for minimum polygon size or by altering the aggregate distance.

Figures 15 & 16 show a comparison between the generated urban tint areas (Figure 15) and the same area from a 1994 USGS 7.5-minute Topographic Map (Figure 16). The generated urban tint reflects very similar boundaries to that of the original Urban Tint. The road generated tint includes areas such as "Brewer Hill" (upper left of maps) and yet exclude areas such as "Suncrest Park" (center of maps). Problems identified with the generation process includes area composition, shape and boundary. These problems are discussed in detail in Section 1.5.1. Additional visualization of the urban tint generated from road data can be seen on VERSION A of the Demonstration Maps.

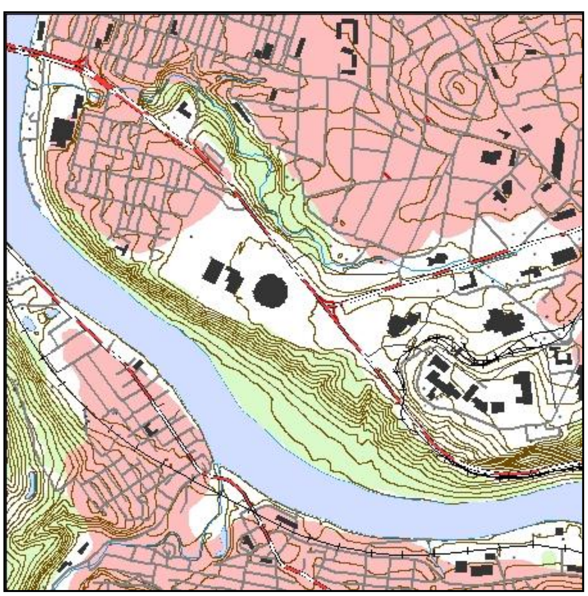

Figure 15 - Urban Tint Generated from Roads

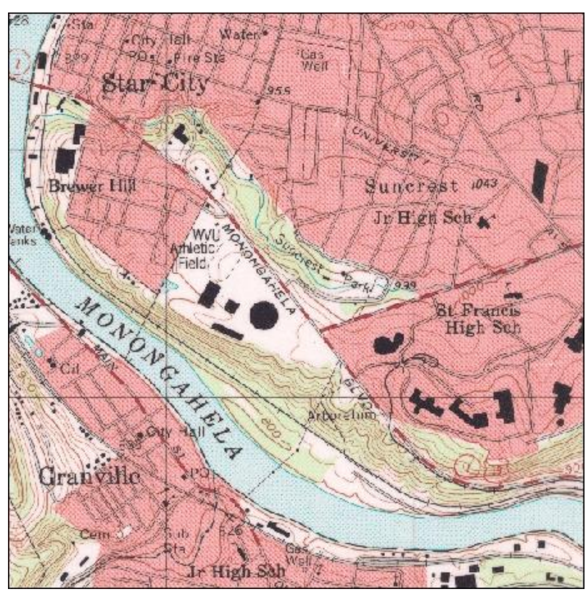

Figure 16 - USGS Topographic map

1.2.0 Urban Tint From Structures

The second approach to generate Urban Tint utilizes structure data consisting of building center points and polygons (Figure 17). The concept behind using structure data to create an urban tint is similar to the use of road data to generate urban tint. A collection of points (structures) can be correlated to the built-up areas and be used to generate the urban tint layer. The process of generating urban tint from structure points is similar to the process used to generate urban tint from road data.

Figure 17 - SAM Structure Point and Polygons

1.2.1 Data

This section of the research primarily used point structure data to identify candidate areas. In addition to the point data, polygon structures were used to enhance the structure point data. The structure data was collected in 2003 as part of the WV Statewide Addressing and Mapping program. All structures were collected with a center point. In addition to the points, building footprints (polygons) were also collected for structures greater than 7500sq ft. Figure 17 shows the source data on the NAIP 2007 aerial photography.

It is recognized that not all states currently have structure point data and fewer areas will have building polygon data. However, structure datasets are recognized as a key data layer and it is anticipated that structure data will become widely available as more states develop their GIS data infrastructure. For this research, both point data and polygon data was used to create the urban tint, although this process can be used effectively with or without polygon data.

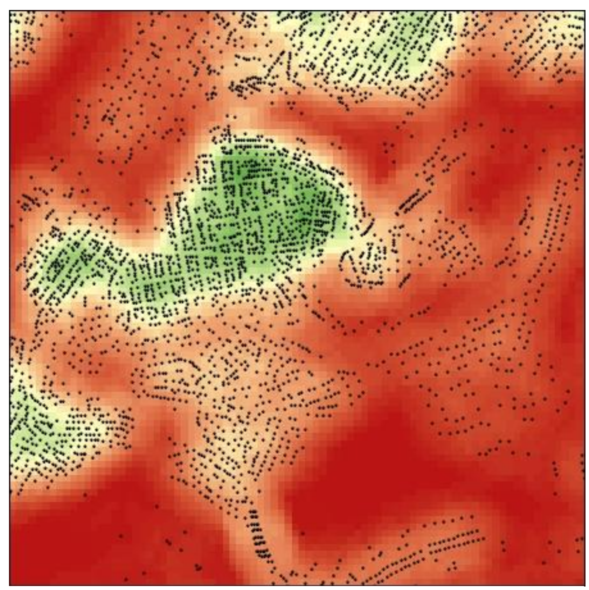

Figure 18 - Density Raster from Structure Points

1.2.2 Density

The point density process analyzed the relationship, between the point data (structures) and how many occur in a given area(statistical neighborhood), and results in a density raster, as shown in Figure 18. We used the USGS NMP Part 6 (Box 1) and directly translated that criteria to GIS, utilizing mathematics to model built-up areas.

> *Density of building and associated use areas will vary from densely concentrated areas downtown, to moderately concentrated residential areas where most of the property is developed.*

Box 4 - Part 6, NMP Technical Instructions

In order to use points to create a density raster, we made assumptions about how much area the point represents. In a residential area that is considered built-up, a house sits on a lot. This lot is considered to meet the conditions of the NMP Part 6 (See Box 4). As this definition explains, the concentration of structures can vary within defined built-up areas. Residential areas would have a single structure surrounded by an open space or yard, which would be represented by points spaced at a distance. Theoretically, residential areas would be less dense than urban core areas. In an urban core area, structures may be adjacent to each other, represented as points spaced closely together.

However, Figures 19-24 demonstrate that this generalization may not always be the case. These figures show NAIP aerial photography on the top row, and their corresponding density grids on the bottom row. All images show the source data structure points for reference. Figure 20 outlines two areas for a comparison between an urban area (upper left outline) and a residential area (lower right outline). Figure 23 shows the un-weighted point density raster of the same area. The color saturation represents density; red colors are low density and green colors are high density. Figure 19 shows the urban core area and the resulting density map (Figure 22). In contrast, Figure 21 shows a residential area that is more dense (Figure 24) than the outlined urban area. The point density for these two areas do not reflect the intuitive and expected result; that an urban area will have a higher density than a residential area. There are multiple factors that determine why the urban point density is lower. As

Figure 19 - Urban Points

Figure 20 - Comparison Sites

Figure 21 - Residential Points

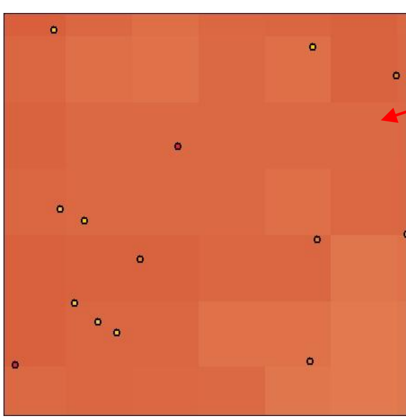

Figure 22 - Urban Density

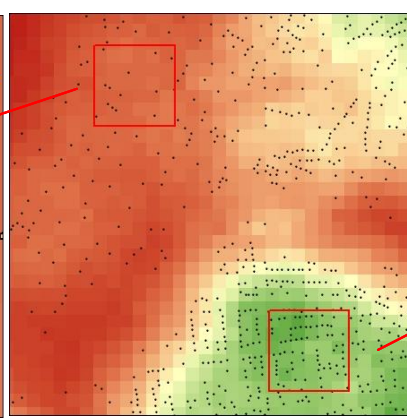

Figure 23 - Comparison Site Density

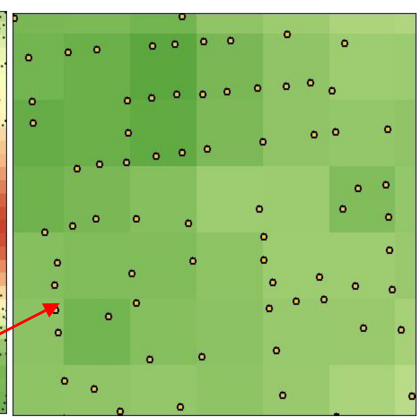

Figure 24 - Residential Density

seen in the images, the building size of the urban structures are much larger than that of the residential areas. In addition, the associated use areas of the urban structures, such as parking lots, are also larger than that of residential use areas (such as yards). This information suggests that point data will need to be weighted in order to better define built-up areas .

Even with the variability of density between urban and residential areas, the designation of an urban tint area can be rendered down to a minimum density threshold. Urban and residential areas that meet this minimum threshold would be candidate for inclusion in the urban tint, while areas that fall below the minimum threshold would not be included. Using a minimum threshold based on an un-weighted density grid will have varying results. Adding a point weight value was one solution used to correct for the density disparity between urban and residential points, and the difference between structure sizes.

1.2.2.a Point Weight by Structure Size

Similar to the road weighting, we looked at assigning weight values to the building points before they are used to create the density raster. The need for weighting the point data became evident after initial tests showed that certain areas populated with large buildings were being under-represented (Figure 25). For example, two points that represent a residential house of 2,000 sq ft and office building of 100,000 sq ft have the same effect on the density raster. The need behind assigning weight values to particular point features reflected the disparity between the density of large structures in the urban core and generally smaller residential structures. Urban core areas often have very large buildings, yet they can be adjacent (high density) or spaced further apart (low density). Strip malls or similar office complexes often contain large structures, parking lots and associated use areas, but because their representative points are spaced further apart, they result in a lower density. This poses a problem when defining the area that would normally be considered built-up.

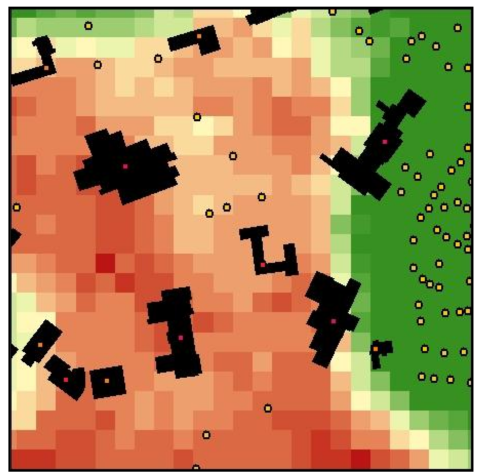

Figure 25 - Density Un-weighted Points

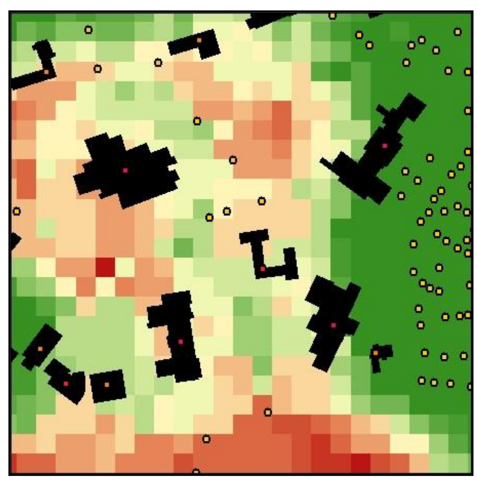

Figure 26 - Density Weighted Points

Assigning a weight value to a point allows it to represent a higher point density (Figure 26) than if it were not weighted (Figure 25). We used a weight scheme that correlated the size of the building footprint with the average value of a single point. An average point represents a structure (house and garage) and its associated use area (yard and driveway), which in total represent 7,500 sq ft of built-up area. This number was slightly less than what the US Census reports as the average lot size in a Metropolitan Statistical Area (8,526 sq ft) sold in 2008. This also works out to be between 1/8 and 1/4 of an acre. Points which represents 7,500sq ft assumed an un-weighted base value of 1. For structures larger than 7,500 sq ft, a weight value was assigned that equaled 2.5 times their measured square feet value divided by 7,500. This formula directly relates the building footprint size value to the average point value. The 2.5 multiplication factor is intended to account for the use area (parking lots) that the building may have, which is not directly reported in the building's footprint size.

1.2.2.b Point Weight Alternative Value

Ultimately, the assigned weight value to a point is a means to better relate the point to the actual structure and use area it represents. In this research, we established an average point value and then assigned higher values to selected points based on structure size. This provides a way to differentiate between points that would otherwise be valued the same interpolated through the density raster. However, this type of weighting can be achieved without knowing the precise size of the building. The principal is the same, regardless of which criteria is used. Zoning, deed or tax information or simply a building type could all allow a similar type of point weight.

For example, zoning could be used as a criteria to classify building points. At a minimum, it could be used to distinguish between residential, commercial and industrial buildings. For example, residential points may represent 7,500 sq ft, commercial points may represent 25,000sq ft and industrial points 50,000sq ft. This basic classification may be enough to differentiate how the point relates to the actual structure and associated use area. In relation to the density map, the residential areas would be held to a higher threshold than commercial or industrial areas. These types of adjustments can be made according to region of the country or data availability.

In addition to zoning, some structure datasets may be attributed with a structure type, such as "residential", "warehouse" or "retail". Almost any type of classification could be applied and used to establish a weight scheme according to the availability of the dataset. More detailed classification of weight values assigned will produce a point density raster that more accurately reflects the real world.

1.2.3 Focal Statistics

Unlike the road data, the point data did not need to be processed through focal statistics. This is in part attributable to the geometry type. It is easier to extract clusters from the point data density raster than from the road density raster.

1.2.4 Reclassify

The resulting density raster is then reclassified into two data classes. This reclassification step acts as a threshold to determine which areas of density are included as candidate areas and which are not. The reclassification plays an important role in determining the size of the overall candidate areas. Altering this threshold can expand or shrink the candidate areas. The product of this reclassification step is a raster which is then converted to a polygon dataset. For this research, our classification threshold value was 2. This threshold was chosen because it represented the best balance of inclusion versus exclusion. After several tests, it was found to be beneficial to be more inclusive during this step. Other processes used later can be used to exclude areas.

1.2.5 Cartographic Processes

After the classification step, both the road generated and structure generated tint rasters undergo the same process. Section 1.4.0 describes in further detail the processes used to clean and enhance the data to complete the urban tint layer.

1.2.6 Research Summary

This research demonstrates that it is a feasible to generate urban tint areas from structure point data. There are drawbacks to using un-weighted point data. However, when weighted building point data or polygon data was used for enhancement, the process had very positive results. Most generated areas were congruent with current topographic urban tint areas. Most areas of discrepancy were areas of change, growth or additions. Other areas of differences noted could be attributed to manual interpretation of the original tint layer. Figure 27 shows the generated urban tint and Figure 28 shows the same area of the USGS topographic map.

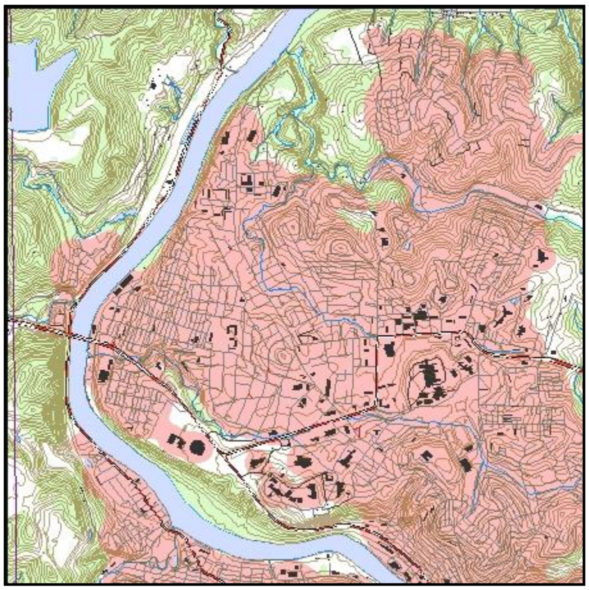

Figure 27 - Urban Tint generated from structures

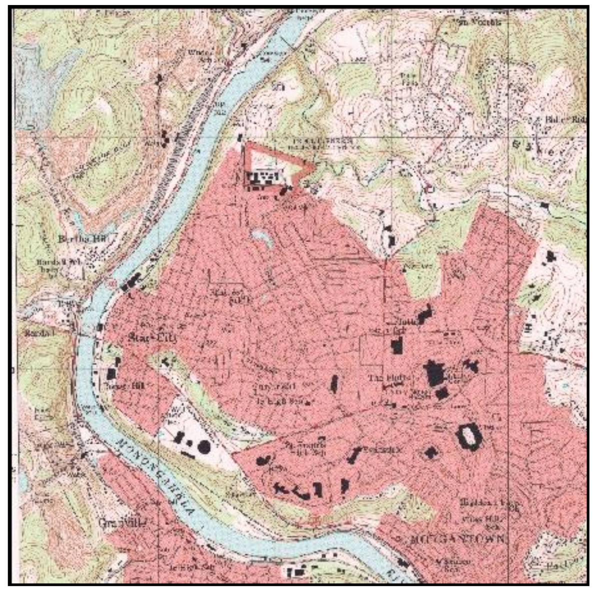

Figure 28 - USGS Topographic Map (Morgantown North)

The urban tint generated from point data has similar problems to the urban tint generated from road data, though to a lesser extent. Of the problems identified with composition, shape and boundary, boundary problems appear to be the most difficult to overcome. Using structure point data can result in a more accurate representation of built-up areas than roads. The point density can be refined at a tighter level than the road data. This is most likely due to the greater variability of point data compared to the road data. In particular, point data density is well suited to identify clusters, while the road data (linear data) density has a tendency of identifying corridors. Using point data can help to eliminate protruding corridors and more closely represent the shape of the built environment. Point density does have unique problems. For example, point density may identify a trailer court as having a higher density than an urban core.

Unfortunately, even weighting has limitations. The point weight value is determined from the size of the building footprint, not the actual size of the building. Therefore the weight value does not distinguish between a one story building (such as a big box store) or that of a multi-story building (as could be found in the urban core). Conceptually, this weighting could use actual square feet if that information is available. Point weighting also tends to over concentrate the actual location of the structure density. This is because a weighted point is only counted in a single cell which it overlaps, it does not distribute density to multiple cells. In reality, the actual building may cover many cells, in addition to the structures use areas. This exhibits itself in the density raster as a rise in density value in the shape of the neighborhood size, centered on the weighted point.

1.3.0 Urban Tint from Roads and Structures

As this report has shown, there are limitations to how the urban tint is generated by both road data and point data. However, one final aspect of creating an urban tint can utilize both road and structure data. This third approach was explored to test the output of a process that uses the structure point and polygon data in conjunction with the road data. There are several approaches that use both road and structure data which produce varying results. For this research, WVGISTC altered the generation process to utilize both road and structure data together. For sake of comparison, WVGISTC also explored two basic GIS processes to combine the individually generated urban tint layers, discussed later in Section 1.3.2.

1.3.1 Process

The overall process for utilizing both road and structure data is similar to their independent processes. The difference is that both datasets were combined during the raster steps using map algebra before a final reclassification threshold was chosen. Data integration earlier in the process allowed both data to better influence the shape of the candidate tint areas. In addition, each dataset can be refined to reflect its strengths during the density weight and classification process steps.

The process begins by individually creating a density raster from each dataset. The two rasters are then individually reclassified. The reclassification gives a common value to both rasters, which are in different density units. The reclassification assigns a higher value to areas of high density and a lower value to areas of low density. The road density is reclassified using Natural Breaks, with four classes while the structure data is classified using Natural Breaks with six classes. This un-balanced classification was chosen to give preference to the structure data, which helps to limit area protrusions introduced by the road data. The two reclassified rasters are then combined using raster math; a function which adds the values of each cells in both rasters and produces a raster with the sum values. The resulting raster is then reclassified again into two classes. For this research, we kept values of 5 and greater, values less than 5 were discarded.

1.3.2 Alternate Tint from Intersect and Merge

For comparison, WVGISTC explored two standard GIS functions that would utilize both road and structure tint layers to create a combined layer. After the urban tints are generated individually, they can either be merged or intersected. The following figures demonstrate the individual product of all three processes for comparison. These figures also display the original boundary of the individual layers. The purple boundary represents tint area generated from road data, orange outlines the tint area from structure data and green outlines the tint generated using both road and structure data.

When the two tint layers are merged, the net result is an increase in area. Figure 33 shows the product of the merge function, which incorporates all areas from both tint layers. Almost as an opposite process, the intersect function uses both tint layers to produce a tint layer containing only areas common to both road and structure tint layers. Figure 34 shows the net result of the intersect function as a subtractive process and produces a much smaller area than either individual layer.

Figure 35 shows the same area, but displays the tint generated using combined road and structure data. The figures show the outlines so that you can see the areas of addition or subtraction clearly. If you compare the green outline, it tends to balance between the purple (road) and orange(structure) areas. It is clear to see that by using both data sources during the generation process, the end product is not purely additive or subtractive. For example; the center of the figures show a cemetery, which has a high road density but does not have structures. By utilizing both datasets for the tint generation, the cemetery area was excluded, as shown in Figure 35.

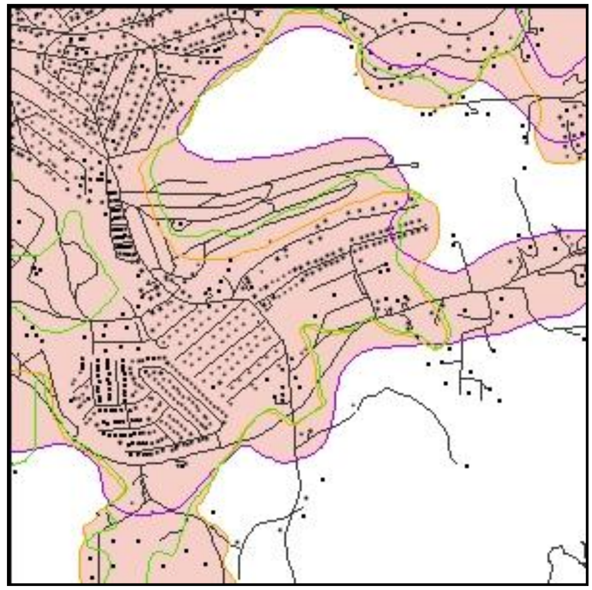

Figure 33 - Merged Road and Structure Tints

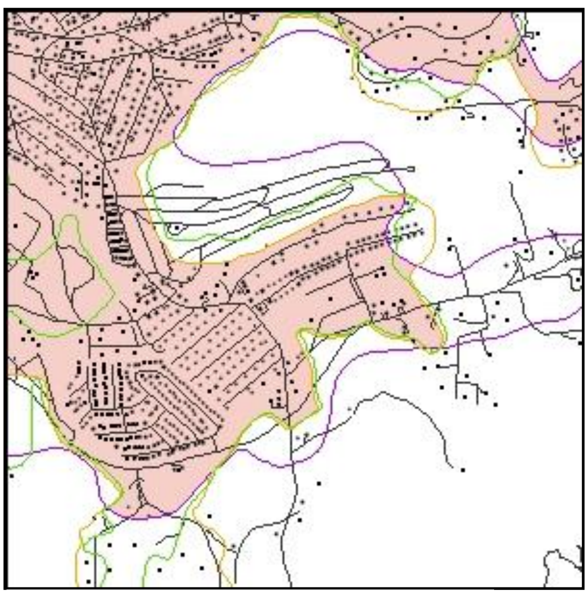

Figure 34 - Intersected Road and Structure Tints

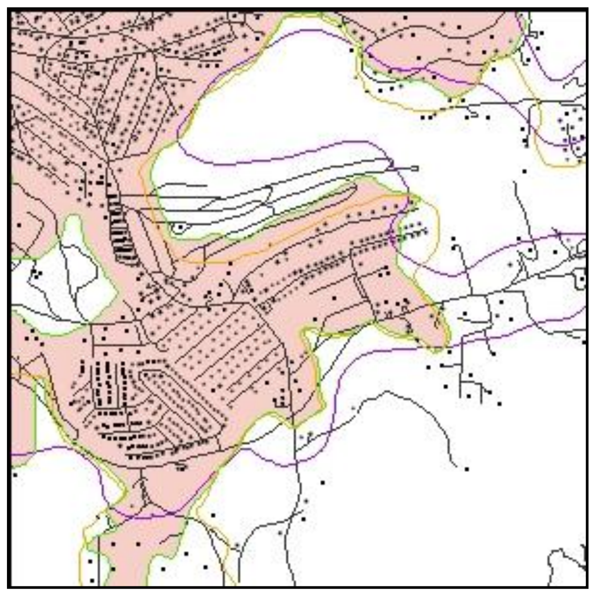

Figure 35 - Generated Tint from Road and Structure Data

1.3.3 Research Summary

Using road and structure data for urban tint generation results in a tint that has a more refined area composition and boundary. While areas that are mutually exclusive in both datasets can still be included in the final data, this approach produces a compromise between the road and structure generated tint layers. The generated urban tint appears to strike a balance between inclusion and exclusion of areas of similar density within their respective datasets. This process allows for both data types to contribute to the final shape and area of the urban tint raster. Initial tests show that by utilizing more information results with a better defined urban tint area. This research has shown that it is feasible to closely match the NMP criteria of using structures and road networks to establish areas that are built-up. These processes can accomplish the task of producing an urban tint based on both structure and road density data. Though the practicality of this process may not be initially viable due to lack of data, as additional data becomes available, this research may be utilized in producing the next-generation of map layers.

Figures 29-32 display the same area of the Lake Lynn Quadrangle at 1:24,000 scale. Figure 29 displays tint generated from road data, which results in the largest tint area. This figure also shows how the tint follows the road corridors on the upper left and top of the map, where the tint continues off the map. Figure 30 shows tint generated from structure data. This figure shows a smaller area than the USGS tint (Figure 32), but shares very similar boundaries overall. Figure 31 shows the tint generated from both road and structure data. This figure shows a similar boundary to the USGS tint area, however has included less area towards the bottom than the structure tint. Figure 32 shows the original USGS topographic map. Though all three generated tints do not match the original USGS tint, they do all convey a close approximation to it. For further visualization of the urban tint generated from both road and structure data see VERSION C of the demonstration maps.

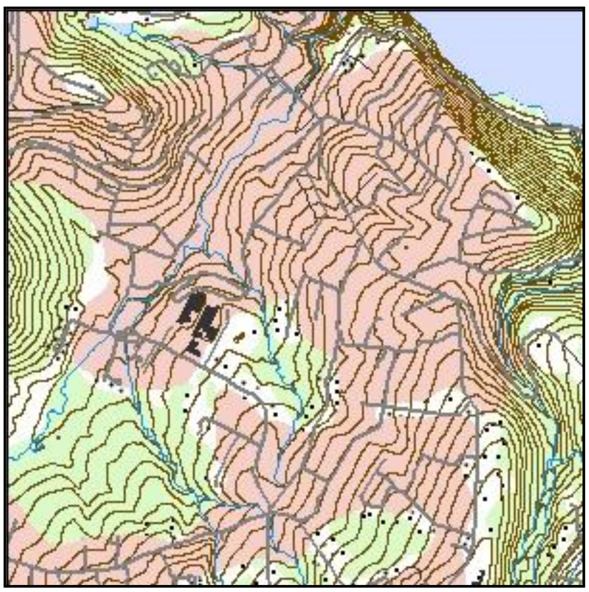

Figure 29 - Generated Tint from Road Data

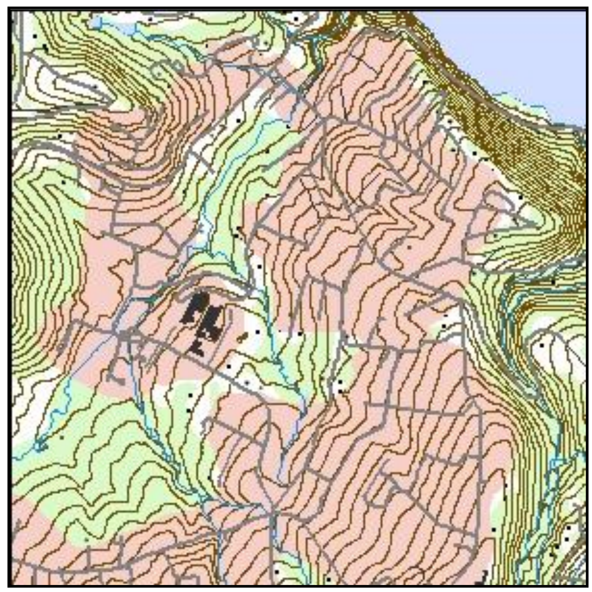

Figure 30 - Generated Tint from Structure Data

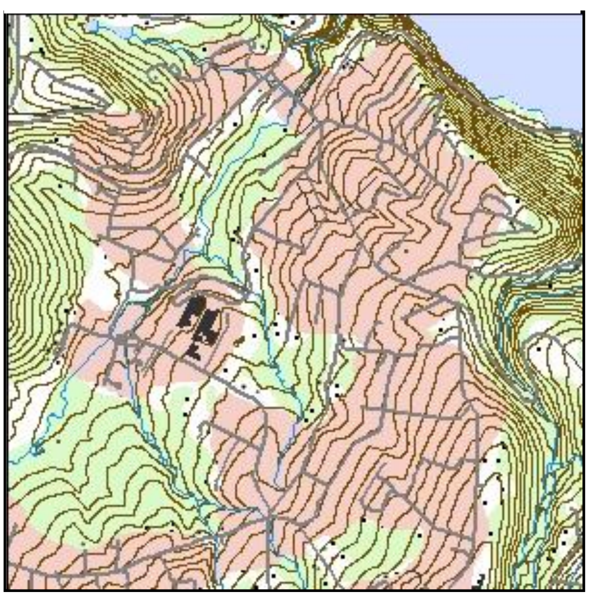

Figure 31 - Generated Tint from Road and Structure Data

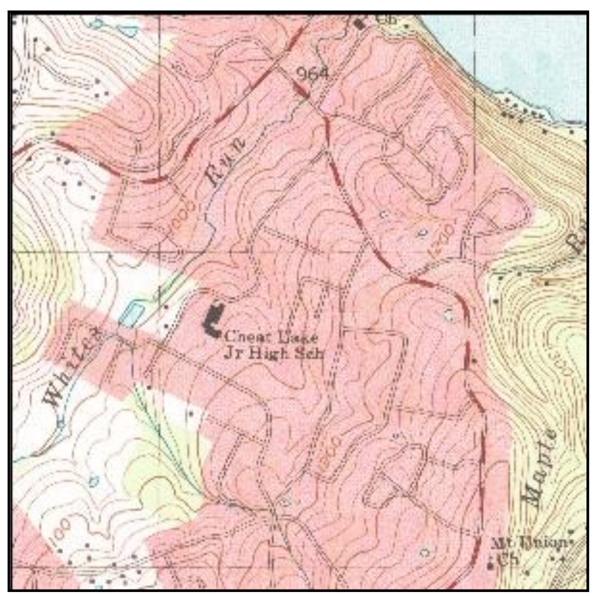

Figure 32 - USGS Topographic Map (Lake Lynn)

1.4.0 Cartographic Process Steps

Although the method of generating candidate urban tint areas may differ, there are similar processes for generating the cartographic output. The cartographic process begins after the density rasters have been reclassified into a two class, integer raster. This reclassified raster is then converted to polygons for further refinement. Any areas of variance from standard procedures will be noted when appropriate.

1.4.1 Raster to Polygon

This step converts the reclassified raster data into a polygon layer. The conversion to a polygon layer is necessary for a number of reasons. Many cartographic enhancement processes and end use of the tint layer require a vector dataset, not a raster data format. For this research we converted the raster to the Shapefile format, though a geodatabase feature class could be used. This step used the reclassified values to create a polygon layer attributed on class value. During this step, we remove all polygons of class value of 1 (See Figure 13). One variable to note using the ArcGIS tool "Raster to Polygon" is the option to simplify polygon features. This option is not used in this step to allow for a separate smooth step later.

> *Capture Conditions: If BUILT-UP AREA is > 30% built-up, and is > 40 acres in size, and is > 660 feet along the shortest axis, then capture.*

Box 5 - Part 6, NMP Technical Instructions

1.4.2 Remove Polygons

This step removes polygons using a threshold based on the size of the polygon. This step is based on the NMP Part 6 (Box 5) minimum area size. Figure 36 shows areas over 40 acres in purple and areas under 40 acres in red.

This step was performed in two different orders during the process; before and after aggregation. Changing the order of this process step has consequences that effect the composition of the candidate areas. See section 1.4.3 for further explanation.

1.4.3 Aggregation

This step performs an aggregation function on the candidate urban area polygons. This step is based on the Part 6 Technical Specification (See Box 6). The aggregation step accomplishes several tasks; first it combines adjacent polygons that are within 660 feet of each other.

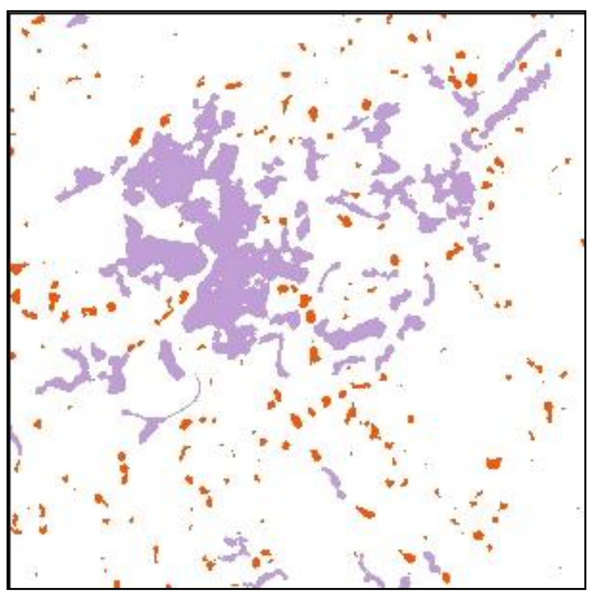

Figure 36 - Candidate areas < 40 acres in red.

This step also removes holes of certain sizes from within polygons. For this research we set a criteria to fill holes less than 3,000 acres. There did not appear to be corresponding USGS documentation regarding the need for a continuous urban tint. Using existing quads as examples, this step seemed appropriate.

Finally, this step removes polygons under a certain threshold based on area size. Our threshold was set at a minimum of 40 acres to correspond to the capture conditions. This part works in conjunction with the previous step, regardless if polygons under 40 acres were already removed.

A large BUILT-UP AREA that is separated from a smaller isolated BUILT-UP AREA, may be extended to include the smaller area, if the smaller area is within 660 feet of the larger BUILT-UP AREA.

Box 6 - Part 6, NMP Technical Instructions

As noted in section 1.4.2, removing undersized polygons prior to the aggregation step is thought to adhere better to the current USGS Technical Specifications. The NMP Part 6 (Box 6) refers to extending larger built-up areas to incorporate smaller built-up areas. This was interpreted to exclude areas that would fall below 40 acres independent of the larger built-up area. By removing these areas prior to aggregation, they will not be included in the urban tint area. When candidate areas < 40 acres are removed before aggregation, the overall remaining polygons will generally have a smaller number of inclusions and therefore cover a smaller overall area.

For this research, the removal of areas < 40 acres was performed before aggregation. Figure 37 shows areas < 40 acres outlined in red and the resulting tint when they are not removed prior to aggregation. This shows the additional areas of inclusion that have been aggregated into larger polygon areas. Figure 38 shows the same extent; areas < 40 acres are shown in red, and the resulting area when they are removed before aggregation. Removing polygons prior to the aggregation step may initially seem positive, but there are drawbacks to this. For example, there are areas that should be incorporated to generalize the boundary, yet are excluded. There are also multiple groups of smaller areas that may total greater than 40 acres when collected as a whole; these areas may traditionally be collected. However, if they individually fall below 40 acres, they are removed and therefore excluded from incorporation. This problem is mitigated somewhat by the focal statistics process (for the road tint), but as noted, this process must balance between allowing islands and overt smoothing of the boundary.

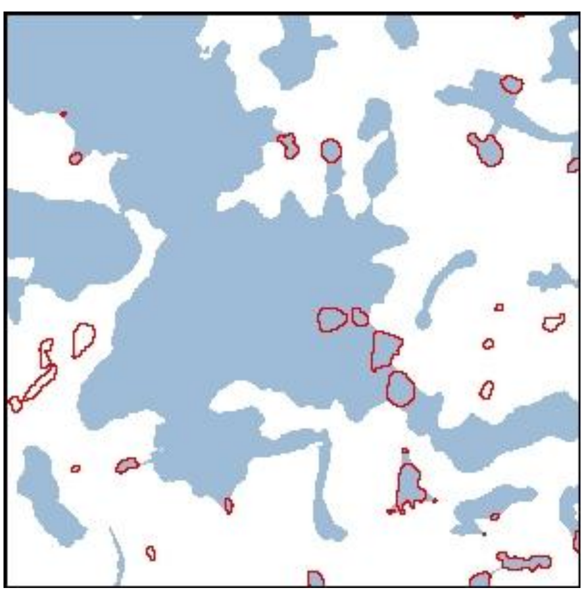

Figure 37 - Areas < 40 acres included in aggregation

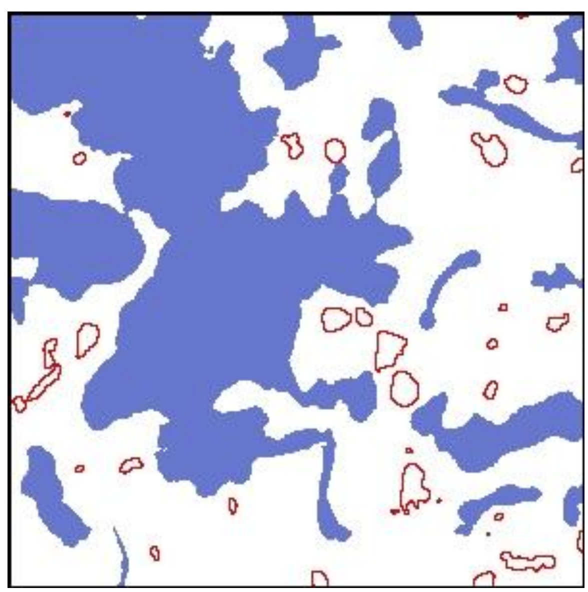

Figure 38 - Areas < 40 acres removed prior to aggregation

1.4.4 Smooth

The step removes the pixel edges and smoothes the polygon boundary for a cartographic quality layer. Figure 39 shows the polygon before processing; note the pixilated boundary edge. Figure 40 shows the smoothed edge and Figure 41 shows the same area in context with other map layers. The variable of smooth tolerance is a balance between an over-simplification of the polygon edge and inclusion of detail, such as the end of a road. For this research, we used a smooth tolerance of 120 meters for both road and structure tint layers.

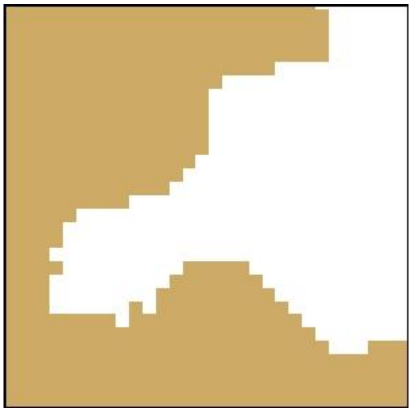

Figure 39 - Polygon Pre-smooth

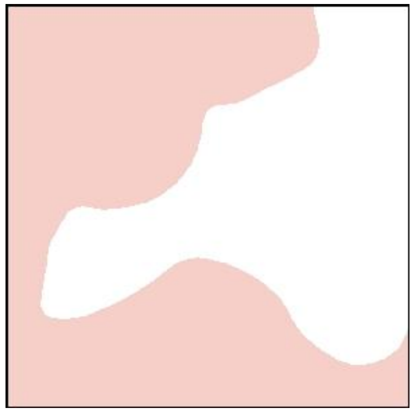

Figure 40 - Polygon Smooth

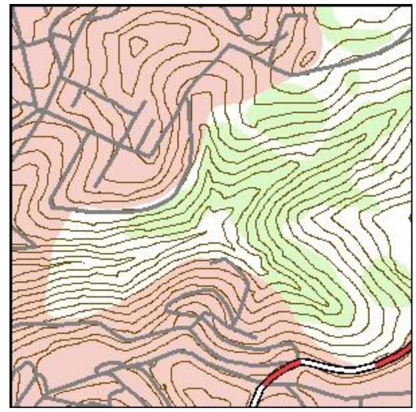

Figure 41 - Generated Tint

1.4.5 Clean / Erase

The final process step cleans the urban tint polygons of certain areas such as rivers, lakes and parks from the generated urban tint layer. This step is to comply with NMP Part 6 (Box 7) criteria for areas not to collect as urban tint. This step is necessary for several reasons. The density raster used to create the candidate areas ignores specific boundaries such as rivers and lakes and the resulting tint polygon may overlap these areas. In addition, small areas such as a city park may be filled in during the aggregate step. The clean process uses the erase tool to subtract undesirable areas of overlap from the final tint layer. All of these layers may not be available as polygon data for use as an erase layer, however larger areas such as rivers and lakes should be widely available. Additional data can be used at any time after this step to perform this step and enhance the urban tints cartographic properties.

1.4.6 Technical Review

Ideally, the generated tint polygons would undergo final inspection from a technician. This inspection might compare new areas with older areas, may clean or adjust certain polygon boundaries and check against any blatant errors introduced through automation. The result of this step would be the final Urban Tint polygon for use in a digital mapping environment.

If the following features meet capture conditions for that feature, then do not collect BUILT-UP AREA:
Lake/Pond
Reservoir
2-D Stream/River
2-D Canal/Ditch
Swamp/Marsh
Railway Yard
Park
Athletic Field (other than tracks)
Cemetery
Golf Course
Exhibition Ground
Outdoor Theater
Runway/Apron/Taxiway

Box 7 - Part 6, NMP Technical Instructions

1.5.0 Final Conclusions

The WVGISTC investigated the use of structure data and road data as separate means to generate an Urban Tint layer. This report has discussed the results of this research and demonstrated what can be achieved using digital methods. This report details the process steps and decisions used to generate the Urban Tint, as well as their potentially positive and negative results. It is intended that USGS will use this research to help initiate implementation of some of these tools and processes in its digital mapping products. The methodology using a weighted road dataset to generate urban tint is currently the most usable approach, given the wide spread availability of road data. However, problems remain with a fully automated process which may hinder the implementation of these digital processes without a review step by a technician. Using a weighted structure dataset appears to produce a more accurate urban tint that is ready for immediate use than an urban tint generated from roads. However, current availability of structure data may pose a significant limitation for most areas. This may change as more structure data becomes available, making the structure process and combined data process more feasible.

1.5.1 Challenges of Automation

This research has identified several key areas of difference between the digital process and the original manual process. The issues that may pose problems to use are related to composition, shape, and boundary. The tint layer composition refers to which areas are included or excluded during the creation of the urban tint layer. Shape refers to the overall boundary (or shape) that the urban tint layer has (such as circular, amorphous or elongated). The boundary issues refer to the large scale, high detail edge that the tint area has and generally how it may perform cartographically. The following sections describe the problems in greater detail, using Figures 42-45 to demonstrate example areas. The figures shown include all three generated tint boundary outlines. The outline colors are: Purple Boundary Tint from Roads, Orange Boundary Tint from Structures and Green Boundary Tint from Road and Structures.

1.5.1.a Composition

In Figure 42, a major area of difference can be seen around "St. Francis High School" (middle right of map). USGS may have originally included this area due to the presence of multiple large structures in the context of a large built-up area. The buildings are on the campus of West Virginia University and are connected by a single road. This type of interpretation would be evident from aerial photography, but not through the automated process. This area was not included in the road generated tint due to a low density of roads. The tint from structures did include the area, but the boundary is much more irregular than the USGS tint.

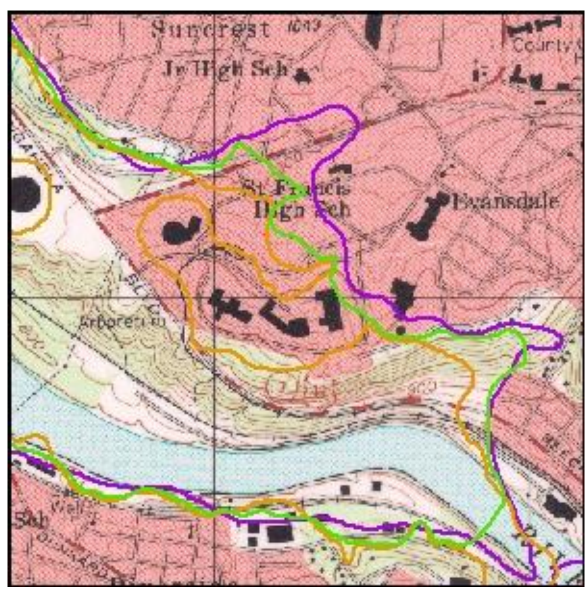

Figure 42 - Generated Tint Outlines and USGS Topographic Map

Although the generated tints can closely approximate the built-up conditions, it is not able to catch all possible conditions that might equate to the original capture conditions. Tints generated from roads were susceptible to misclassified roads which can incorrectly identify candidate areas. This was shown on Version A of the Lake Lynn Quad, which incorporated two sections of incomplete highways that were classified as local roads. Other composition differences can result from the traditional method of manual interpretation which identified structures, roads and use areas in photography. Using the road and structure data alone to determine density may excluded areas which should be included. For example, Figure 19 showed few points in an urban environment, leading to a low density. However, the large parking garages and parking lots were not represented by the structure data.

This problem could be overcome by using additional data, such as an enhanced weight scheme or additional data to enhance or verify built-up areas. One possibility is to check identified candidate areas against land cover data for verification. This is an area that needs further research.

1.5.1.b Shape

> *Narrow strips where there is only a road and buildings along both sides of the road should not be included even if they would connect an adjoining larger area that meets capture conditions.*

Box 8 - Part 6, NMP Technical Instructions

The generated tint area generally matches up well with existing USGS urban tint areas. However, the automated method has a tendency to include protruding road corridors that lead into and out of the urban area. This is an undesirable trait as described in the NMP Part 6 (Box 8). These candidate areas often reflect the urban build up surrounding larger urban areas, specifically road corridors adjacent to, but detached from larger urban areas. This problem is more prominent in the Urban Tint generated from roads, but the Urban Tint generated from structures also showed this problem in a few areas as well.

Figure 43 shows the generated tint area outlines on top of a USGS topographic map (faded for clarity of boundary lines). This graphic demonstrates how protruding areas are included in larger areas. The figure shows that the tint area boundaries all follow a road corridor out to a larger built-up area. All areas within the outlines meet the minimum threshold for creating an urban area and are therefore included.

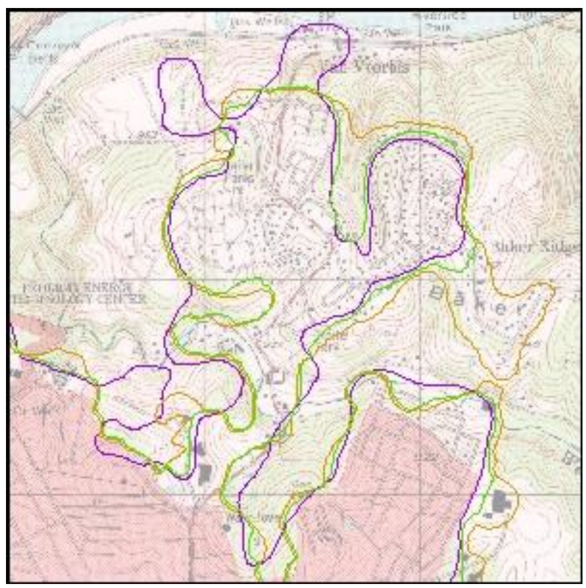

Figure 43 - Generated Tint Outlines and USGS Topographic Map

The generation process must strike a balance between being inclusive of an urban areas edge, without extending the boundary of the main area which includes areas along a corridor. This is difficult based solely on density. Therefore the shape of the candidate areas must factor into the final urban tint area. Utilizing shape of candidate polygons was explored by using a Minimum Bounding Rectangle script, but not fully researched in this study. This is discussed further in Section 1.5.3.c.

Another possible solution could be to use algorithms that identify the core area of the candidate areas, this information could then identify protruding areas outside the core area, which could be excluded from the final Urban Tint polygons or marked for trimmed. This step could be included in the automation process, however using any automated process to make these distinctions may leave an undesirable or abrupt boundary. This is an area for further research.

> *Although Built-Up Area limits need not be fixed in relation to linear features, where practical, the limits of Built-Up Area should share the edge with other linear features such as roads, railways, stream/rivers, boundary lines, etc.*

Box 9 - Part 6, NMP Technical Instructions

1.5.1.c Boundary

The third issue identified is the structure of the edge or boundary of the generated urban tint. Boundary delineations are ignored during the generation of urban tint. The NMP Part 6 (Box 9) states that the edge of urban tint areas should follow and adhere to logical boundaries such as roads, streams or other edge features. The automated process is not able to consider such cartographic determinations when the shape of the polygon is constructed. Figure 44 shows how the USGS urban tint has been constrained to a municipal boundary to the right and bottom, and constrained to a road towards the top. The generation process is not able to reflect these types of cartographic determinations. The three tint boundaries have correctly identified this area as built-up, but are not able to adhere to the strict boundary as cleanly as the original USGS boundary.

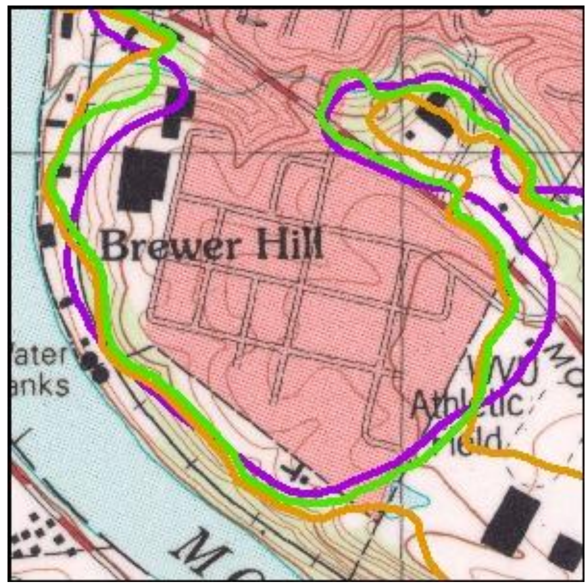

Figure 44 - Generated Tint Outlines and USGS Topographic Map

The generated tint boundary may also overlap areas that would seem illogical, such as a river. The generated urban tint can overlap these areas for several reasons; the source data may not accurately reflect a built-up condition. An example, of this may be a bridge (density over water) or a cemetery (high road density, no structures). Another reason may be the density raster or focal statistics rasters were classified or smoothed, which may cause high density areas to cover non-built-up areas. A third possibility is that the aggregate function may have incorporated an area that does not meet the built up criteria, such as shown in Figure 45. The areas on both sides of the river meet the capture conditions, and they are within 660 feet (the aggregation distance) and therefore are

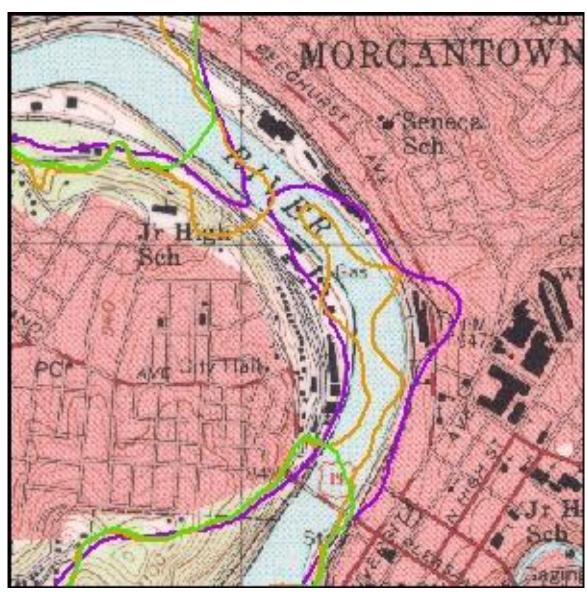

Figure 45 - Generated Tint Outlines and USGS Topographic Map

aggregated together to cross the river. The boundary of the generated urban tint areas show the areas of greatest difference between the manual USGS process and this digital process. The edge differences may also be the most challenging of these issues to overcome through automated means. One possible solution could be to use topological tools that can alter the polygon boundary to snap to adjacent features. This would require additional data sources, such as municipal boundaries or streams, but this should not be a limiting factor. Ultimately the tint areas could meet the NMP guidelines by fully automated methods, but further research in this areas is needed.

1.5.2 Research Findings

The USGS National Mapping Program Standards set forth in the Part 6 manual were intended to create USGS Urban Tint areas as cartographic features for printed topographic maps. The original intent of Urban Tint areas was to denote areas of dense, built-up areas and to signify generalization of building structure symbology. Those same standards set to a digital process can result in areas that do not reflect the spirit of the original urban tint area. However, this research has shown that the USGS process and criteria used can be altered to create areas that more closely reflect the spirit of the original urban tint, if not the exact areas. There are several areas which we have identified as candidate areas of change, which may help transition traditional USGS methods towards automated digital methods.

1.5.2.a Transitioning Standards

There are difficulties in producing a clean, cartographic boundary to the tint area. However, the edge of the tint area may be less of an issue in a digital environment than on paper. Digital mapping techniques such as scale dependency or data generalization may allow for the urban area to be displayed with good cartographic principals. Print maps generated at 1:24,000 scale can still use the newly generated urban tint, though the urban tint edge may not be as refined as the manual process. These new urban tint polygons can be used to identify areas of high density, and therefore be used to identify areas that may need to have generalized building symbology. The generated tint polygons can be used to clip building point datasets for use as a scale dependant layer.

Preliminary research for an improved raster using both structure point and road data as an enhanced integrated study to identify built-up areas resulted in improved urban tints. Though this approach may be feasible in the future, it may be less practical now as structure data is not as widely available as road data.

1.5.2.b Tint Generation Hybrid Approach

The process outlined in this report does present challenges; however, the majority of the automated process results in a product which stays accurate to the goals of the project. One potential solution is a compromise to both traditional cartographic techniques and use of automation. By using the processes outlined in this research, USGS could generate new candidate areas as data sources become available. Once these urban tint polygons are generated, there could be follow up from a technician using digital editing tools to clean, or adjust the boundary where necessary. This could still save considerable time from the manual process of identifying the urban tints. It is conceivable that the automated process could be run on an annual basis or target designated areas for updating. This could be done as needed or from a local National Map partner as data becomes available or workloads permit.

1.5.2.c Utilization

The urban tint that results from this process may have purpose in a digital mapping environment, without manual processing. At moderate scales (ex. 1:24,000), the urban tint can be used to mask or clip structure point data. This would allow the map reader a clearer look at the data and fulfills the original spirit of the urban tint. Figures 46-48 demonstrates such an example. Figures 46-47 are digitally generated maps displayed at 1:24,000 scale. Figure 46 shows an area of high structure density without use of an urban tint mask, displaying all structures. Figure 47 shows the same area with the generated tint polygon (un-edited) used as a building mask. For reference, Figure 48 shows a mosaic of the original USGS Topographic maps. When compared to Figure 48, the generated urban tint on Figure 47 does not match the original tint exactly. However, when compared to Figure 46, it does perform the same task as was intended from the original urban tint. For instance, though these two maps do not display any road annotation, it is quite clear which map would be able to better accommodate road feature annotation.

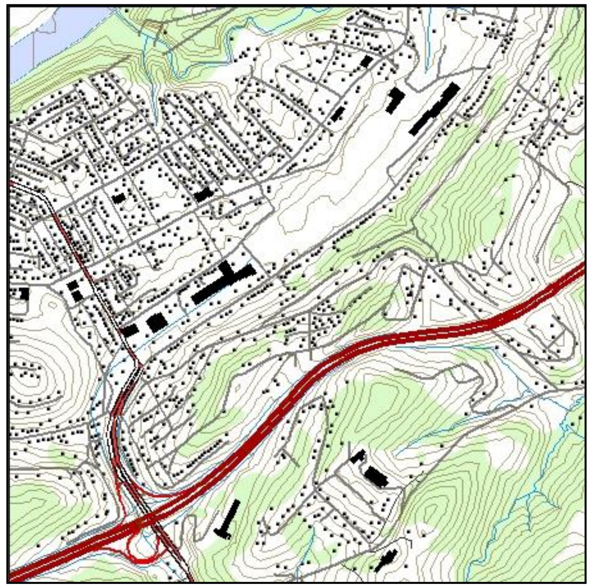

Figure 46 - Generated Topographic Map

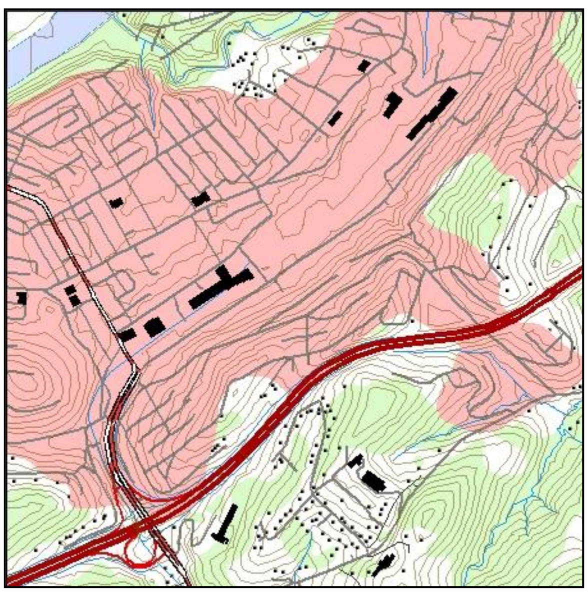

Figure 47 - Generated Urban Tint and Topographic Map

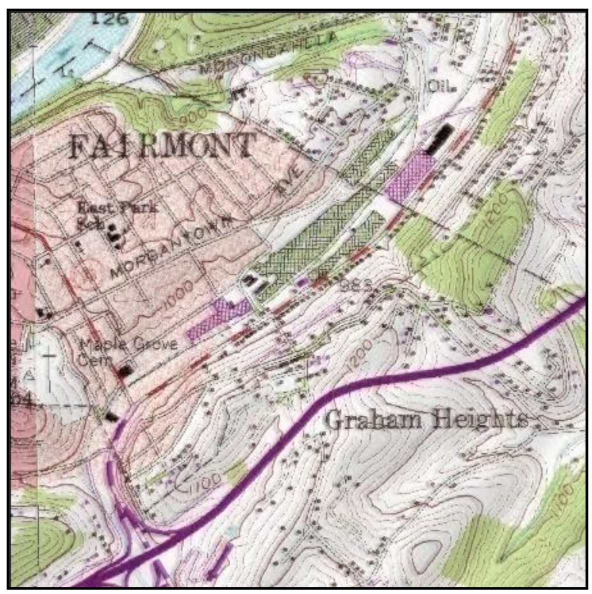

Figure 48 - USGS Topographic Map Mosaic

1.5.3 Further Research

This research has tested numerous variables that can affect the output of the generation process. However, many additional tools and concepts remain untested. Additional processing steps may be able to help reduce the problem issues or enhance the output of the generation process. Some of these areas for further research are listed in this section.

1.5.3.a Automation Script Development

This research performed the tasks outlined in this report through manual means. Automated scripts could be developed to allow for a fully automated process, allowing large input datasets to be processed by USGS as they become available. This automation could be used as an update tool when new data becomes available. The scripting could be written using Python or Visual Basic and potentially operate as a custom tool in ArcMap or Arc Catalog.

1.5.3.b Improved Boundary

One of the issues identified in this research is the boundary or edge of the generated urban tint. This problem was not addressed through this research, but new research using custom tools or topology may be able to improve or replicate the desired cartographic principals of the original urban tint.

1.5.3.c Improved Composition

As the automation process is developed, additional tint criteria could be used for the selection of candidate areas. One such criterion could be based upon the shape of the area. Figure 49 shows one such process which uses a script to identify the minimum bounding rectangle of the area. The minimum bounding rectangle information can then be used to measure the overall width and length of the area and give each polygon a length to width ratio. Using this technique, it may be possible to identify corridors or elongated areas that are undesirable and remove them before they are included in the final tint area. Initial exploration of this technique appears to be promising, but the script as written only considers the full width and length of the bounding rectangle, not the actual polygon. Further research is needed.

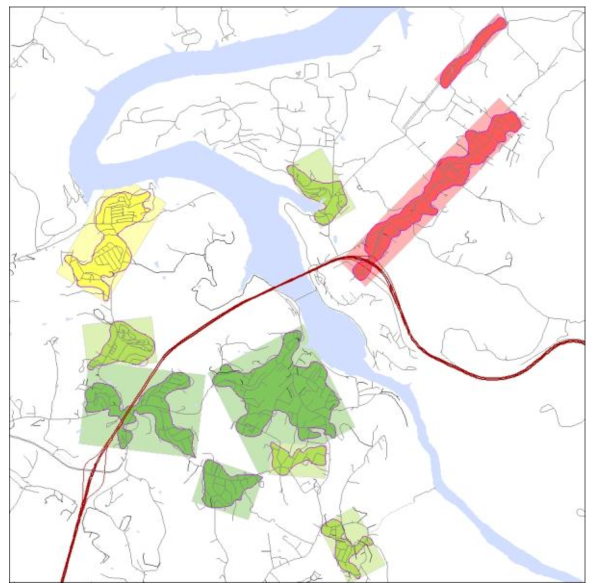

Figure 49 -Polygons shown by Length/Width

1.6.0 Demonstration Topographic Maps

The demonstration topographic maps are meant to show the results achievable using digital processes for Urban Tint Generation. There are four variations of three USGS quadrangles: Morgantown North WV-PA, Morgantown South WV and Lake Lynn WV-PA. Versions A, B and C each display the Urban Tint layer as generated from various data sources. Version A shows the Urban Tint generated from road data. Version B shows the Urban Tint from Structure Data (both point and polygon). Version C shows the Urban Tint resulting from using both road and structure data. Version D shows no urban tint; rather it is meant to display all the road and structure data used to generate the Urban Tints. There was a simplified process and single data source for the forest tint generation which resulted in one tint layer, which is displayed on all three versions. See Section 2.0.0 for further information on the Forest Tint Layer.

Section 2
Forest Cover Tint Generation

2.0.0 Forest Tint from NLCD

The green tint symbology that appears on the USGS 7.5 minute topographic maps is intended to represent tree cover (Woodland). The NMP Part 9 (Box 1) describes the Woodland tint. This data has traditionally been collected under photogrammetric standards that specify requirements of canopy height, minimum widths and tree density which result in a green shaded area (polygon) that symbolizes woodlands. USGS documentation refers to "Trees"," Woods" and "Woodland" and for the purpose of this paper all are considered the same. This section researched the method and data source to provide for a suitable digital feature that can be used with digital mapping products.

> *TREES - An extensive area of land covered with woody perennial plants having a self-supporting main stem or trunk, a definite crown and usually reaching a mature height of more than 6 feet.*

Box 1 - Part 9, NMP Technical Instructions

The data most suitable for creating a forest cover tint using digital processes is the National Land Cover Dataset (NLCD). Three classifications of the NLCD can be used to represent the forest cover tint. This section of research discusses the process used to extract the useful data and create the tree cover tint layer.

2.1.0 Data Source

The NLCD is created and supplied by a federal program with well defined standards and periodic, standardized updates. The NLCD is published by the Multi-Resolution Land Characteristics Consortium (MRLC), which is a group of federal agencies formed to develop a national, uniform, land cover dataset. The NLCD data is consistent across the nation making it well suited as a data source for the generation of a forest cover tint. NLCD is published in a raster format which was created by an unsupervised classification of Landsat data. The NLCD uses a modified Anderson Classification and has a spatial resolution of 30 meters, originating from the Landsat 30 meter cell size.

In addition to forest cover, this data could be used to provide a more detailed tint layer, such as showing distinct coniferous and deciduous forest areas. Though the NLCD may be coarse for some applications, the NLCD could also be used to show orchards, cropland and wetlands where more detailed data is not available.

> *41. Deciduous Forest: Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent 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 percent 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 percent of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.*

Box 2 - MRLC NLCD Classification Description

2.2.0 Process

The WVGISTC used ESRI ArcGIS products for the process of creating the forest cover tint. In summary, the NLCD data is reclassified in to two classes: NLCD Values 41,42,43 (which represent Deciduous, Evergreen and mixed forest types) are assigned a class value of '1' and all other values are assigned a class value of '0'. This reclassified raster is then converted into a polygon, where all '0' values are removed. The remaining polygons then go through a smooth process. For the purposes of this research, these processes were performed manually, however a script could be written to fully automate the process according to the desired workflow. A detailed outline of the process is below.

2.3.0 NLCD Process

2.3.1 Reclassify Raster

This step separates out the usable data from the data not used for creating the new tint. Figure 1 shows the Anderson classified NLCD.

a. Tool: Spatial Analysis>Reclass>Reclassify
    i. Input: National Land Cover Dataset
    ii. Reclassify: Gridcodes 41,42,43 to '1', All other values to '0'
    iii. Output: NLCD Reclassified

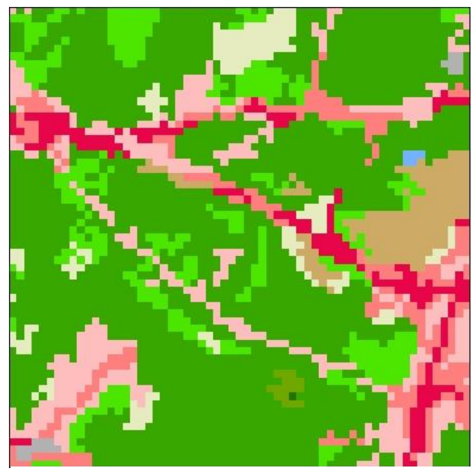

Figure 1 - Classified NLCD Raster

2.3.2 Raster to Polygon

This step converts the reclassified raster into a shapefile polygon. Figure 2 shows the re-classified raster. Purple shows forest areas, teal shows non-forest areas.

b. Tool: Conversion Tools> Raster to Polgon
    i. Input: NLCD Reclassified
    ii. No Simplify polygons
    iii. Output: NLCD Polygon

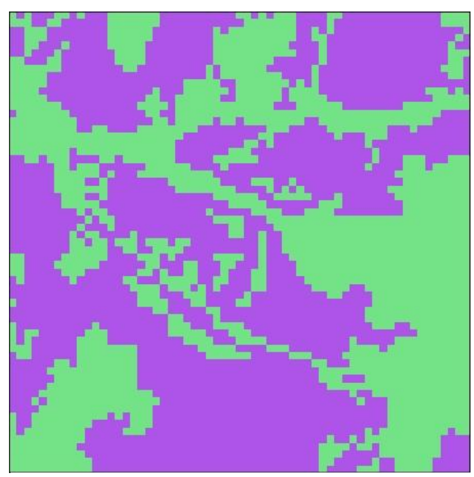

Figure 2 -Reclassified NLCD Raster

2.3.3 Remove Polygons

This step removes the unused non-tint data polygons. Figure 3 shows the remaining forest cover polygons over aerial photography for reference.

c. Input: NLCD Polygon
    i. Process / Tool: Select
        1. [gridcode] = '0'
    ii. Process / Tool: Delete
        1. Delete Selected
d. Output: NLCD Clean

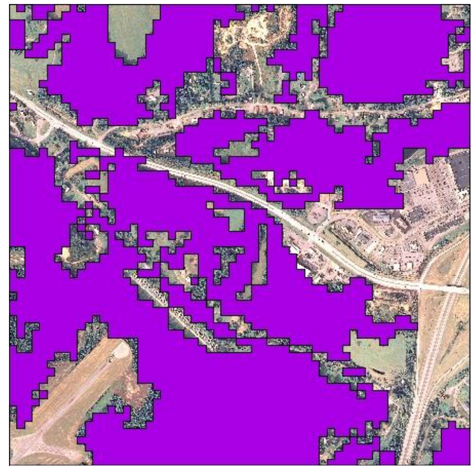

Figure 3 -Extracted Forest Class Polygons

2.3.4 Smooth Polygon

This process smoothes the polygon edges for cartographic use. Figure 4 shows the un-smoothed polygon in purple and smoothed polygon in green.

e. Tool: Data Management Tools > Generalization > Smooth Polygon
    i. Input: NLCD Clean
    ii. Paek: 120 meters
    iii. Preserve endpoints
    iv. No Check
    v. Output: NLCD Smooth

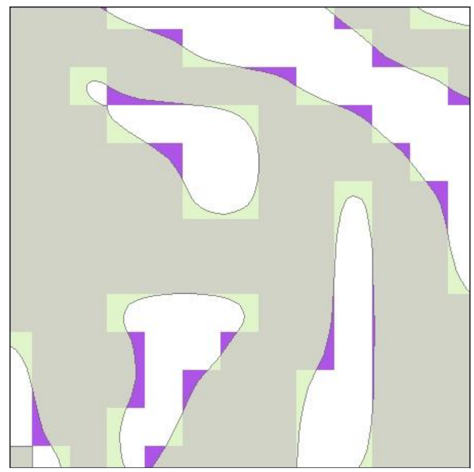

Figure 4 -Original Polygon and Smoothed Tint

2.3.5 Remove Islands (Optional)

This step removes single pixel polygons. In this step, creating a new field to add acreage information allows for selection and removal based on area criteria. This step can be skipped to include all areas. Figure 5 shows a single pixel polygon highlighted in blue outline.

f. Input: NLCD Clean
    i. Add "Area" field
    ii. Calculate "Area"
    iii. Select: Select < .22 acres
    iv. Delete Selected
g. Output: NLCD Clean2

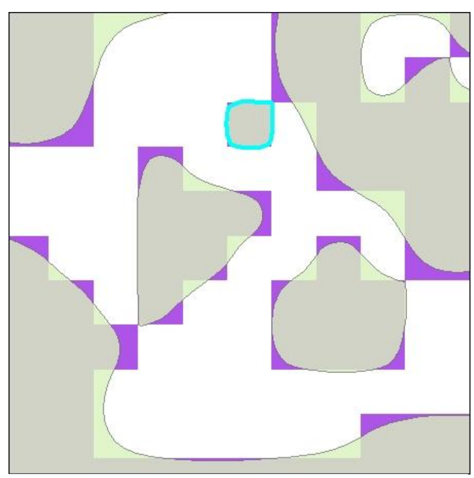

Figure 5 - Single pixel shown in blue outline

2.3.6 Clean

This step removes unwanted areas from the generated tint areas. These areas would follow the NMP Part 9 Tree Cover as areas that are not to be collected as Forest Tint. Figure 6 shows the green tint overlapping a hydrography layer. In this example, these overlaps would be erased.

h. Tool: Analysis Tools>Overlay>Erase
    i. Input: NLCD Clean
    ii. Erase Input: Layer
i. Output: Final Tint

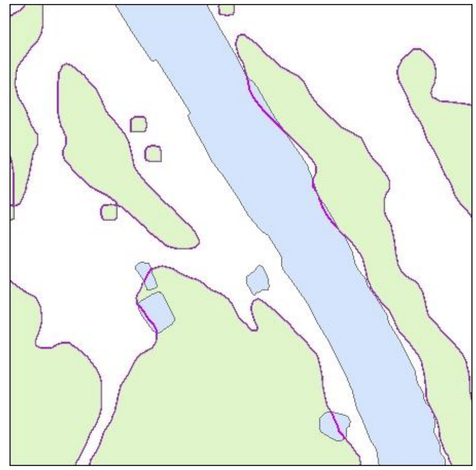

Figure 6 -Tint shown overlapping water

2.4.0 Generation Challenges

The following are some potential issues that arrive from generating a forest cover tint layer from the NLCD

2.4.1 Data Currency

The current NLCD was published in 2004 using Landsat imagery from 2001. The MRLC recognizes the temporal nature of its land cover dataset and is currently working towards updating the NLCD to include 2006 data. It could be assumed that if the NLCD tree cover would change, those changes would likely be areas of subtraction, such as areas of urban growth. Because many areas of subtraction are created and then processed independently of the tint layer, they could then be applied as they are revised to the tint layer. Areas of new tree cover growth would be captured with updates of the NLCD, and therefore would be incorporated into a revised tint as the NLCD is released.

2.4.2 Manual Interpretation

The NMP Standards state the following : "*Depict forested areas that have been clear-cut, but are expected to be reforested, as TREES.*" Though this type of manual decision making has been removed through automation, the problem may be less relevant with dynamic data as it was when creating a paper based product. The update cycle of digital data allows for a more accurate and temporal layer.

2.4.3 Delineation

The NMP Standards state the following: *"The limit of TREES is the extent of the treed area. The limit is the center of the outermost trees, not the edge of their crowns."* This interpretation is not feasible using this methodology. The NLCD 30-meter cell approximates tree density, but does not locate a specific crown or tree. Therefore, the generated edge will be only an approximate tree boundary. This may not be as accurate as a manual interpretation.

2.4.4 Data Resolution

The NLCD source data (Figure 7) is based on 30-meter Landsat Data. It could be assumed that a tint generated from the NLCD directly would then appear pixilated or coarse, as shown in Figure 8. However, with correct processing, this is not necessarily true. The process outlined in this report incorporates a smoothing step to compensate for the pixel edges and produce a cartographic quality layer. Figure 9 shows the resulting tint in context with other layers. Though the pixel edge has been corrected by smoothing, the actual tint boundary may be more of an issue at larger scales. At these detailed scales, the real edge of the tree cover may not match the symbolized tint. However, the forest cover representation is still able to reflect the approximate tree cover location. Figures 7-10 depicts the source data and resulting forest tint products at 1:4,800 scale. Figure 8 shows the source raster edges with NAIP aerial photography shown for reference. Figure 10 shows the USGS Topographic Map for comparison. These figures demonstrate how a smoothed tint area can act as a cartographic quality forest tint.

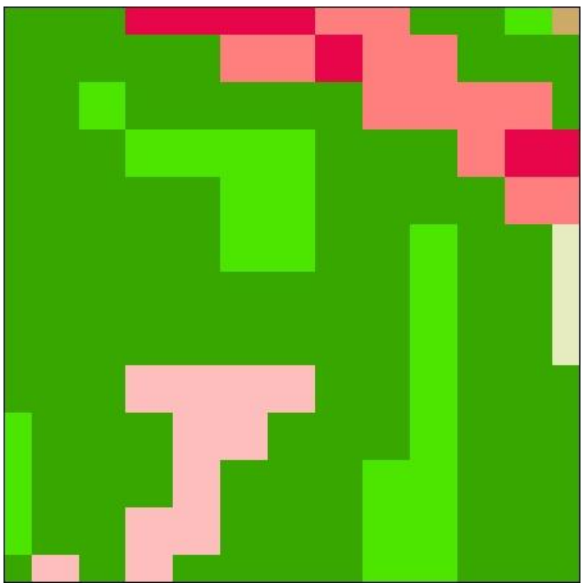

Figure 7 - Classified NLCD Raster

Figure 8 -Forest classes converted to polygon

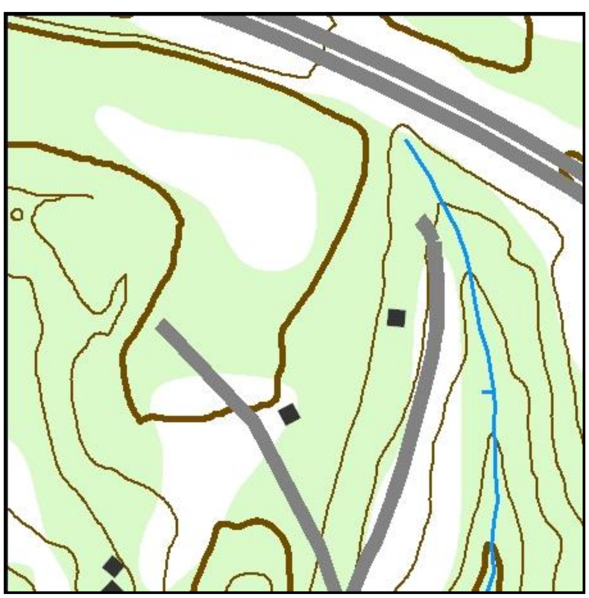

Figure 9 - Generated Tint

Figure 10 - USGS Topographic Map

2.5.0 Research Summary

This research has shown that generating a new forest tint layer using the NLCD has worked exceptionally well. The digital processes outlined in this report could be fully automated, allowing USGS to process large areas. Figures 11-14 show a comparison between USGS Topographic maps and the generated tint at two scales. Figures 11 & 12 displays the maps at 1:10,000 scale. Figure 11 shows the USGS Topographic map. Figure 12 shows the corresponding area using a digitally generated map with the NLCD generated forest tint. Figures 13 & 14 show the maps at 1:24,000 scale. Figure 13 shows the USGS Topographic map, and Figure 14 shows the Generated map and tint. For additional visualization, please see the Morgantown North, Morgantown South and Lake Lynn demonstration maps that use the generated forest tint from this research. There are three versions of each quadrangle, however the forest tint layer is identical on each map.

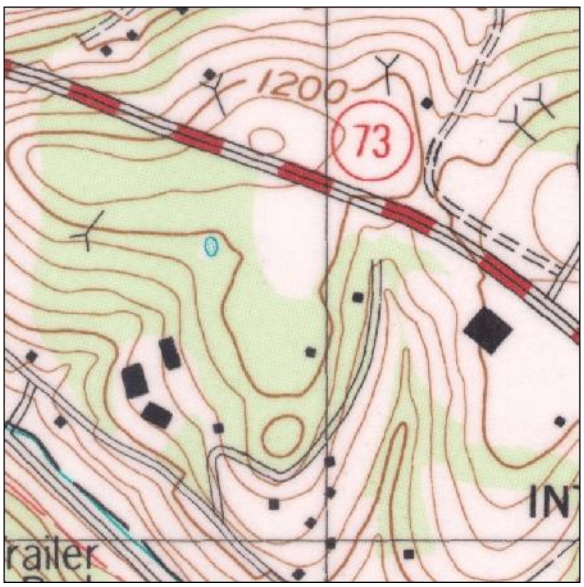

Figure 11 - USGS Topographic Map, 1:10k Scale

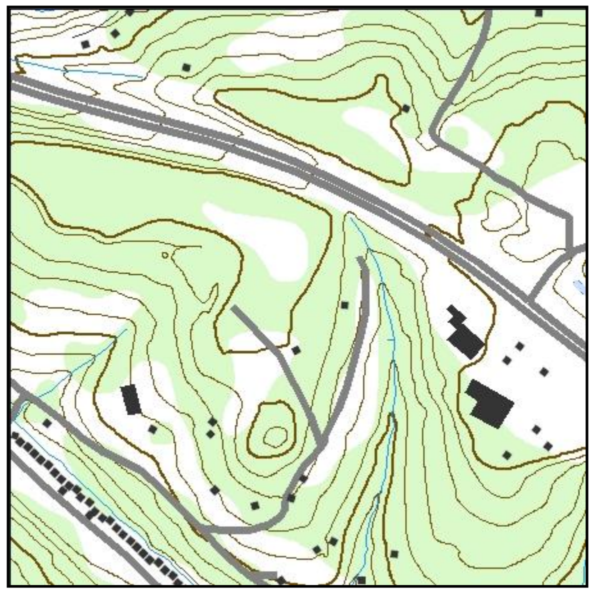

Figure 12 - Generated Tint and Map Layers, 1:10k Scale

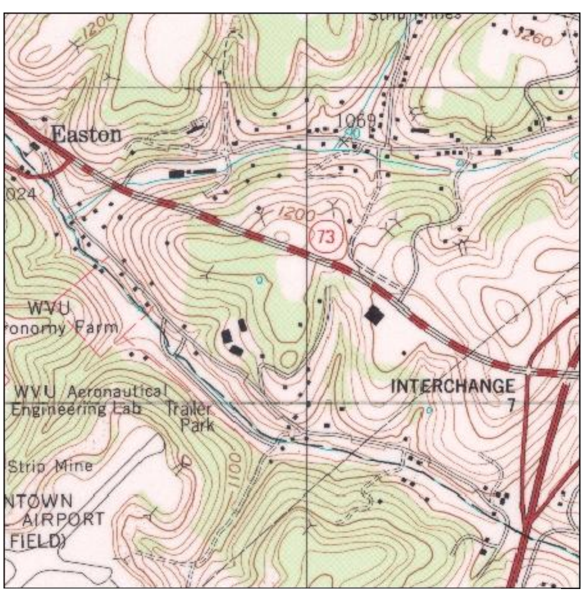

Figure 13 - USGS Topographic Map, 1:24k Scale

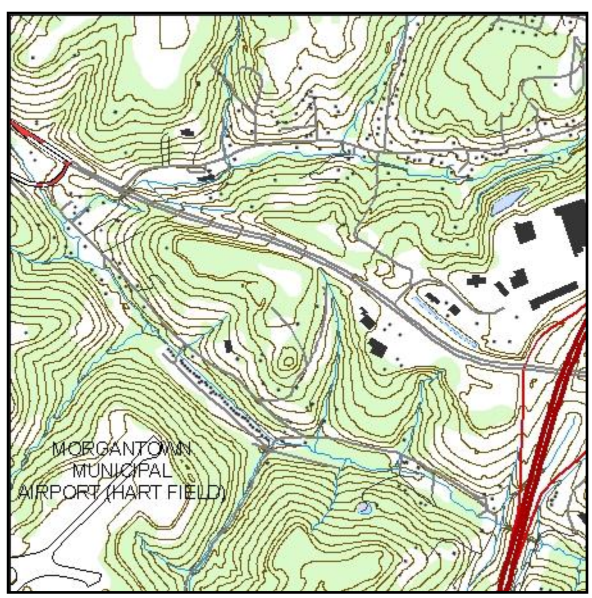

Figure 14 - Generated Tint and Map Layers, 1:24k Scale

Section 3
Structure Symbology

3.0.0 Structure Symbology

This report describes the use of point features and polygon features to represent structure (buildings) features that replicate USGS NMP Part 6 cartographic standards. As structure datasets become more widely available, they are often being utilized on digital map products. USGS uses a distinctive set of symbology to represent structures on 7.5 minute topographic maps. There are two types of generic structure representation: a single, simple square that represents average structures and a more complex shape that represents larger structures. This report is not meant to address specific structure symbols such as churches or schools. There are two brief sections that addresses each of these structure symbologies. Section 3.1.0 discusses the simple shape symbol and how to achieve a point symbol rotation for alignment to road features. Section 3.2.0 discusses the simplification of detailed polygon features for building representation in digital map products. These processes were performed using the ESRI ArcGIS suite of software.

3.1.0 Simple Structure Symbology

This process describes the application of using a point dataset as a representation of structures for use in digital map products. The 7.5min series USGS topographic maps use a 2pt, black, simple block symbol to represent generalized building locations outside urban tint areas (NMP Part 6-20 thru 6-31). For cartographic aesthetics, the block symbols are rotated to align with its corresponding street. These are procedures that replicate the structure orientation.

3.1.1 Structure Point Rotation

This process was done using ESRI ArcGIS – Arc Catalog Toolset using the Near tool. This process requires a structure point dataset and a road dataset at approximately the same scale. The Near tool will calculate the distance and direction to the closest road feature from each individual point feature. This information is written into the point datasets attribute table. The distance will not be used and may be deleted; the Near angle field will be used to specify the rotation angle for the features symbology. When symbolizing the building point dataset, under the advanced properties, there is an option to specify a field that contains the rotation angle. The Near angle field should be specified here, and the option of Arithmetic should be chosen. Figure 1 shows the un-rotated 2pt block symbol. Figure 2 shows the same block symbols but rotated using the Near Angle added through this process.

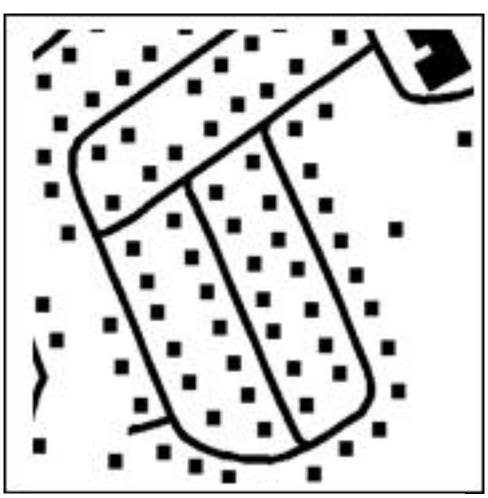
Figure 1- Structures not aligned to roads

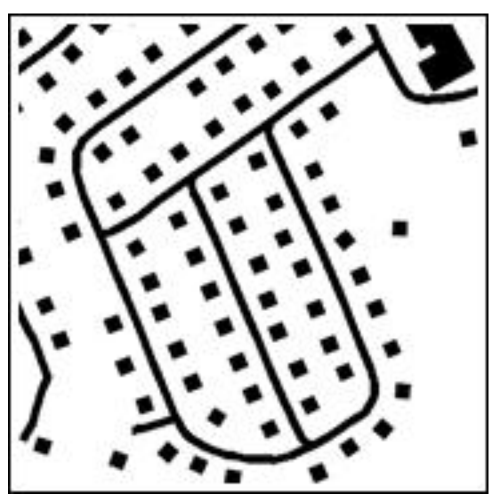
Figure 2- Structures aligned to roads

3.2.0 Complex Structure Symbology

Complex structure polygons can add detail to map products, but they can also make the map un-necessarily complex and difficult to read. In addition, large structure datasets may hinder functionality or performance of online digital map systems. Structure simplification reduces the polygons complexity and can produce a smaller dataset which allows easier readability and greater performance. This section describes the process and tool used to simplify building structures.

3.2.1 Structure Simplification

Within ArcGIS there is a specific tool called "Simplify Building". This tool processes the input polygons and removes certain vertices, resulting in a simplified shape. The tool has two options: the simplification tolerance and minimum area. Testing of the simplification tolerance showed a desired range of 30 feet and 80 feet, depending on the desired level of simplification. Figure 3 shows the original structure polygons. Figure 4 shows the simplified building with a tolerance of 80ft. Figure 5 shows the comparable USGS topographic map for reference.

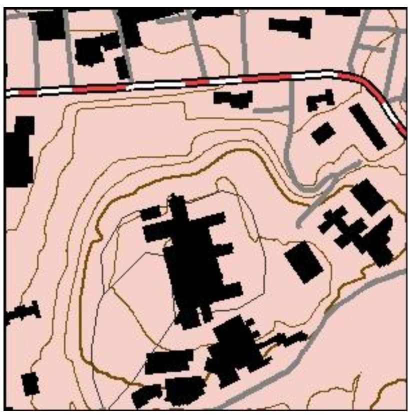

Figure 3 -Original Structure Polygons

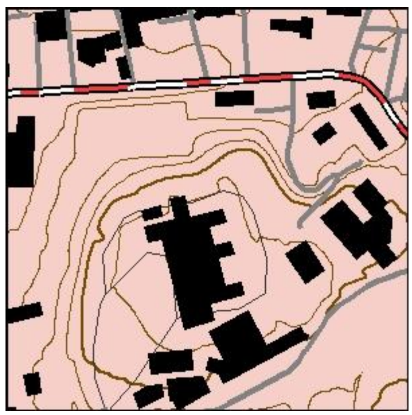

Figure 4 -Simplified Structure Polygons

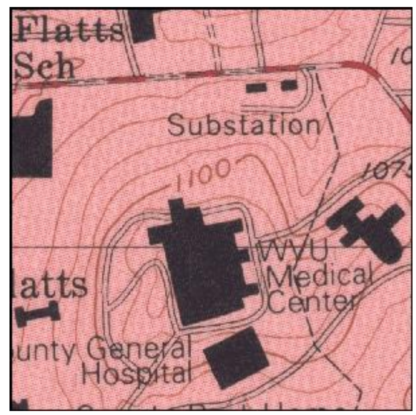

Figure 5 - USGS Topographic Map

Section 4
Appendix - Digital Processes Technical Specifications

4.0.0 Technical Appendix

This section of the report details the process steps, functions, specific tools and their variables that were used in this research. The demonstration maps follow these particular specifications, though certain figures found in this report may reflect different variable values or processes. Input data and output data names have been simplified for sake of clarity. .

4.1.0 Generation Process: Urban Tint from Road Data

1.) Create Input: This step extracts the usable features from the Census TIGER data.
- Process: A series of functions and tools were used to perform this task.
  - Input: US Census TIGER Data
    - Tool/Process: Select if MTFCC = x (x = MTFCC)
      - S1100 (Primary Roads)
      - S1200 (Secondary Roads)
      - S1400 (Local Roads)
      - S1730 (Alley)
      - S1740 (Private Road)
    - Tool/Process: Export Selected
  - Output: Transportation Input

2.) Classification Weight: This step creates a new field in the road data and assigns feature weights using the field calculator.
- Process: A series of functions and tools were used to perform this task.
  - Input: Transportation Input
    - Tool/Process: Add new field
      - Field Name: weight
      - Field Type: double
    - Tool/Process: Select if MTFCC = x
    - Tool/Process: Field Calculator (If x Then = v)
      - x = MTFCC , v = code weight value
    - Calculate Value = v
    - MTFCC code weights:
      - Primary Roads - S1100 = 0
      - Miscellaneous- S1730,1740 = 1
      - Secondary Roads - S1200 = 2
      - Local Roads – S1400 = 3
  - Output: Transportation Weighted

3.) Create Density Raster: This step creates a density raster from the road data using feature weights.
- Tool: Line Density
  - Input: Transportation Weighted
    - Weighted on MTFCC 1-3
      - Population: weight value
      - Output Cell Size:10 meter
      - Search Radius: 120 meter
      - Area units: Acres
  - Output: Road Density

4.) Focal Statistic: This step processes the density raster through a focal statistics process.

Tool: Focal Statistics

Input: Road Density

Statistics Type: Mean

Neighborhood Shape: Annulus

Neighborhood Size: Inner Radius 2 cell, Outer Radius 15 cell

Output: Road Focal

5.) Reclassify: This step reclassifies the focal raster into two integer classes.

Tool: Reclassify

Input: Road Focal

Method: Manual

Class Values: 2 classes, excluding '0' values

Value 1 $\leq$ 1.9

Value 2 > 1.9

Output: Reclassified Road Raster

4.1.1 Generation Process: Urban Tint from Structure Data

1.) Calculate Polygon Area: This step creates a field in the structure polygon data, values are populated by the field calculator.

Process: A series of functions and tools were used to perform this task.

Input: Structure Polygons

Tool/Process: Add new field

Field Name: Area

Field Type: float

Tool/Process: Field Calculator

Field: Area

Calculate Area = Square Feet

Output: Structure Polygons

2.) Intersect: This step assigns the calculated area to points that have a corresponding structure polygon.

Process: A series of functions and tools were used to perform this task.

Tool/Process: Intersect

Input 1: Structure Point

Input 2: Structure Polygon

Output: Structure Point Weight

2.) Create Weight Value: This step assigns the average area (7500) and the point weight value to all points. The calculated area (square feet) divided by 7500 and values > 1 multiplied by 2.5.

Input: Structure Point Weight

Tool/Process: Select

Field Name: area

Select if 'Null'

Tool/Process: Field Calculator

Field: Area

Calculate Selected Features

Values: = 7500

Tool/Process: Field Calculator

Field: Weight

Value : area/7500

Tool/Process: Select

Field Name: weight

Select if > 1

Tool/Process: Field Calculator

Field: weight

Calculate Selected Features

Value: weight x 2.5

3.) Create Density Raster:

Tool: Point Density

Input: Structure Point Weight

Population: weight

Output cell size:30

Search Radius: 30

Area units: Acres

Output: Structure Density

4.) Reclassify:

Tool: Reclassify

Input: Structure Density

Method: Manual

Output: Reclassified

Values: 2 class, excluding '0' values

Value 1 $\leq$ 2.0

Value 2 > 2.0

Output: Reclassify Structure Raster

4.1.2 Generation Process: Urban Tint from Road and Structure Data

1.) Create Road Input: This step creates the classified road data raster.

Complete Section 4.1.0 Steps 1-3

Tool: Reclassify

Input: Road Focal

Method: Natural Breaks (Jenks)

Output: Reclassified

Values: As defined, 4 class, excluding '0' values.

Output: Reclassify Road 2

2.) Create Structure Input: This step creates the classified structure data raster.

Complete Section 4.1.1 Steps

Tool: Reclassify

Input: Structure Density

Method: Natural Breaks

Output: Reclassified

Values: As defined, 6 class, excluding '0' values.

Output: Reclassify Structure2

3.) Add Rasters: This step adds the classified structure and road data rasters and produces a combined raster.

Tool: Plus

Input 1: Reclassify Structure2

Input 2: Reclassify Road 2

Output: Combined

4.) Reclassify:

Tool: Reclassify

Input: Combined

Method: Manual

Output: Reclassified

Values: 2 class, excluding '0' values

Value 1 ≤ 4

Value 2 > 5

Output: Reclassify Combined

5.) Cartographic Process: Complete 4.1.3

4.1.3 Cartographic Process

1.) Raster to Polygon: This step converts the reclassified raster to a polygon dataset.

Tool: Raster to Polygon

Input: Reclassify

Output: Road/Structure Polygon

2.) Delete Polygons

Process: A series of functions and tools were used to perform this task.

Input: Road/Structure Polygon

Tool/Process: Select Features

Field Name: Grid

Select if 1

Tool/Process: Delete Selected

Tool/Process: Add new field

Field Name: Area

Field Type: float

Tool/Process: Field Calculator

Value = Calculate Acres

Tool/Process: Select Features

Field Name: Area

Select if < 40

Tool/Process: Delete Selected

Output: Road/Structure Less40Acres

3.) Aggregate Polygon

Input: Less40Acres

Output: Aggregate

Aggregation Distance: 660 feet

Min Area: 40 acres

Min Hole Size: 3000 acres

4.) Smooth Polygon

Tool: Smooth Polygon

Input: Aggregate

Smooth Algorithm: Paek

Smoothing Tolerance: 120 meter

Output: Urban Tint

4.2.0 Simple Structure Symbology Process

1.) Add Rotation Angle: This adds the necessary rotation angle.

Tool: Near

Input: Structure Point Data

Near Feature: Roads

Search Radius:1000 ft

Option: Include Angle

Output: Structure Point Data

2.) Rotate Angle: This is done in ArcMap.

Symbology/Advanced/Rotation

Field: Near_Angle

Rotation Style: Arithmetic

4.2.1 Simplification Process

1.) Simplify Structure: This tool simplifies the structure polygon.

Tool: Simplify Building

Input: Structure Polygon Data

Tolerance: 80 feet

Simplification Tolerance: (none)

Output: Structure Polygon Data