ABSTRACT

Research has shown that the amount of impervious surface in a watershed is a reliable indicator of the impacts of development on water resources. While attention has focused on quantifying this relationship, little work has been done to assess the efficacy of various methods for mapping impervious surface area. This paper compares spatial representations of percent imperviousness against photogrammetrically-derived calibration and validation data from high resolution digital planimetric datasets for 52 towns in Connecticut and New York. Impervious surface estimation techniques examined include: (1) NLCD 2001 impervious surface layer, derived through regression tree classification of Landsat ETM data, (2) a population density and land cover-based regression model, using US Census Bureau population and NLCD 2001 land cover, and (3) land use-specific coefficients for NLCD 2001, as modeled with the Impervious Surface Analysis Tool (ISAT). In comparing results with the reference data, it was found that estimates of tract-wide imperviousness based solely on land cover coefficients (ISAT), while an easily implemented method, yielded the lowest accuracy with an RMSE of 5.65. The direct Landsat-based method (NLCD IS) showed an RMSE of 5.48. The population density and land cover-based regression model, which leveraged readily available input information, demonstrated the highest accuracy of the three techniques studied with an RMSE of 4.56. The impervious surface estimation methods summarized should provide decision makers and planners with the information to guide them in selecting the optimal method of mapping imperviousness given their programmatic needs and technical resources.

INTRODUCTION

Nonpoint source pollution (NPS) is a top contributor to water quality problems in the United States (EPA, 1994). A major cause of NPS is urbanization that increases the volume, duration, and intensity of stormwater runoff (Booth et al., 1993). Past research has found that the amount of urban runoff and its impact on stream conditions and water quality in a particular analysis unit (census block, tract, watershed, etc.) are strongly correlated with the percent area of impervious surface (IS) within that unit (Schueler, 1994; Arnold et al., 1996; Clausen et al., 2003). This makes the amount of IS an important indicator of water quality (Schueler, 1994; Herlihy et al., 1998; Brabec et al., 2002; Boyer et al., 2002; Clausen et al., 2003; Roy et al., 2003). This also mandates the need to develop accurate, consistent, cost-effective, and replicable techniques to measure impervious surfaces at a variety of scales. To be viable, such methods have to make use of publicly available land cover and other ancillary data. This paper reports on several approaches to estimating percent imperviousness and is part of ongoing research aimed at developing a suite of analysis tools for effective land management (Civco et al., 2002; Wilson et al., 2003).

EFFECTS OF IMPERVIOUSNESS

Impervious surfaces (IS) are surfaces that prevent the percolation of water into the soil and the evapotranspiration of soil moisture and ground water to the atmosphere. As is common, we only consider IS as related to anthropogenic development. With this, there are two major groups of features that can be called impervious: rooftops (buildings, pools, and patios) and transportation system (roads, sidewalks, driveways, and parking lots).
Imperviousness has a direct effect on local streams (creeks, rivers, etc.) as well as indirect effects on downstream receiving waters (ponds, lakes, etc.) (Schueler, 2003). Among the affected characteristics of a given watershed are hydrological (the amount of runoff, peak discharge rates, and baseflow), physical (stream morphology and temperature), water-quality (nutrient and pollutant loads), and biological (stream biodiversity) (Arnold and Gibbons, 1996; Brabec et al., 2002).

The relationship between the amount of IS in a watershed and the watershed’s hydrological parameters has been well documented (Booth and Jackson, 1997; Schueler, 2003). Generally, an increase in the amount of IS leads to greater stormwater runoff and peak discharge rate as well as a decrease in groundwater recharge (Gilbert and Clausen, 2006; Jennings and Jarnagin, 2000). The result is increased storm stream flow rates (and possibly flooding) and lower summer base flows.

Increased imperviousness also impacts the geometry of the stream network and its temperature. As noted in Klein (1979), higher runoff flows erode stream banks causing the stream channel to become wider and straighter, possibly destabilizing the stream and destroying riparian habitats. On the other hand, lack of proper groundwater recharge leads to lower water depths during dry periods, resulting in more rapid warming in the summer and cooling in the winter (Brabec et al., 2002). Water temperature increase during the summer can also be caused by runoff from heated IS (Van Buren, 2000; James and Thompson, 1997). This negatively affects fish and plant populations.

IS, by its nature, serves as a depository for numerous atmospheric and man-made pollutants. During rain events, these are transported by stormwater to downstream aquatic systems. Transportation-system related contaminants (oil, grease, de-icers, MTBE, etc.) wash into public water supply in potentially harmful concentrations (Oberts, 1986). They tend to accumulate in sediment that continue to release toxins even in the absence of the original source of pollution (Mason et al., 1999). Likewise, trace metals (zinc, copper, lead, and others) may enter watersheds through transportation system runoff (Brattebo and Booth, 2003; Gilbert and Clausen, 2006). They also may be delivered to IS (and subsequently to water bodies) by atmospheric precipitation as well as by runoff from rooftops and painted structures (Chang et al., 2004; Davis et al., 2001). The main concerns associated with these metals are toxicity and possible carcinogenic effects.

The amount of potentially pathogenic microorganisms in surface waters has been found to be positively correlated with the degree of development in the watershed (and thus the amount of IS) (James and Thompson, 1997). Associating the presence of E. Coli and other Coliform bacteria in surface waters with IS may serve as an additional indicator of the impact of IS with water quality.

Another concern is the IS-related nutrient intake. Nutrients, in particular phosphorus and nitrogen, come from urban stormwater contaminated with fertilizer and organic waste, as well as from atmospheric deposition (Boyer et al., 2002). Their presence in streams affects the growth rate of algae and may lead to eutrophication (Carpenter, 1998).

The hydrological, physical, and water-quality effects of IS can have significant impact on aquatic animal and plant populations (May et al., 1997; Booth and Reinelt, 1993). Specifically, stream organisms may not be able to adjust to temperature changes and simplified geometry of affected streams (Miltner et al., 2004). Increased stream velocities and flow volumes during storms uproot river vegetation and endanger spawning habitats (Schueler, 2003). On the other hand, abrupt changes in water levels due to decreased groundwater recharge lead to fish overcrowding. Toxin, pathogen, and nutrient pollution also have adverse effect on biodiversity.

**STUDY AREA**

Ten towns in the state of Connecticut and 42 towns in Westchester County, New York served as the study sites for calibrating and validating the impervious surface estimation models (Figure 1). The sample included rural, suburban, and urban towns.
DATA

All Connecticut and New York datasets used in the study were first reprojected into the State Plane NAD83 coordinate system for the respective state. Planimetric data portraying the built-up landscape – the photogrammetrically-derived layers delineating building footprints, roads, driveways, parking lots, and other anthropogenic impervious surfaces (Figure 2) – served as validation data for each of the methods examined, and as calibration for all but one – the NLCD 2001 IS dataset, which was developed independently of this project.

Connecticut planimetric data were obtained from the appropriate municipal or county GIS departments. The towns of Groton, Milford, Stonington and Suffield, were dated 2002. The other town planimetric data, created from older aerial imagery were updated via on-screen digitizing using high resolution ADS-40 true-color coastal imagery and/or Connecticut 2004 digital orthophotographs (Figure 3).

All Connecticut planimetric datasets were in ESRI shapefile format and contained polygons assigned to a single impervious class, regardless of the original impervious feature class (e.g. building, driveway, parking lot, etc.).

New York planimetric datasets, circa 2000, were obtained from the Westchester County GIS department. They consisted of two IS layers for each town: structure (buildings) and transportation (roads, sidewalks, driveways, and parking lots). For the purpose of this study, the two layers were combined using the ArcGIS “Union” command.

National Land Cover Data (NLCD) were obtained from the United States Geologic Survey (USGS). The dataset was extracted from springtime leaf-off and summertime leaf-on Landsat ETM. It was chosen because of the data’s nationwide availability, thus enabling possible geographic extension of the models under consideration to parts of the country other than the Northeast. Once processed, this raster dataset had a resolution of 100 by 100 feet per grid cell (Figure 4).
Figure 2. Planimetric impervious surface data for the town of West Hartford, CT

Figure 3. Example of initial and updated planimetric data, updated using aerial digital orthoimagery circa 2004
Census 2000 tract boundaries for Connecticut and New York were obtained from the TIGER (Topologically Integrated Geographic Encoding and Referencing) files from the Cartographic Boundaries section of the U.S. Census Bureau website. There were a total of 303 tracts used in the study. Census tract boundaries included many positional inaccuracies and were subject to considerable editing. They were adjusted to match road centerlines, municipal boundaries, and water body shorelines as depicted on the municipal planimetric data and the available town boundary information. These corrections were necessary to ensure that tracts could be overlain accurately with planimetric and other digital datasets. To examine the relationship between land cover, population density, and IS, significant water areas bordering the towns in the study were removed from datasets. This editing applied especially to those municipalities bordering Long Island Sound and/or having a large river bounding the municipality. Population density (in people per square mile) was calculated from the Census 2000 population tables using the edited area of each tract within the study municipalities.

METHODS

The methods for estimating percent imperviousness included: a general Classification and Regression Tree (CART) NLCD sub-pixel analysis; the Impervious Surface Analysis Tool (ISAT) with NLCD land cover-specific coefficients; and a regression model incorporating both the NLCD land cover data and population density data. A description of each follows.

NLCD sub-pixel analysis

As part of the NLCD 2001 program, along with land cover and forest canopy closure, estimates of percent imperviousness were developed (Yang et al., 2003). Landsat ETM+ data and derived Tasseled Cap transform, along with ancillary data including elevation, slope, and a soil index, were used in a general CART algorithm to produce rule-based models for prediction of continuous measures of imperviousness. Yang et al. report an average error of predicted versus actual percent of IS from 8.8 to 11.4% for three test areas – Sioux Falls, SD, Richmond, VA, and the Chesapeake Bay area. A sample of the NLCD sub-pixel impervious surface estimate is shown in Figure 5.

---

1 These data were not derived from this study. They are product of the NLCD 2001 program.
NLCD 2001 land cover

NLCD 2001 sub-pixel IS estimate

Figure 5. Examples of NLCD 2001 land cover and sub-pixel IS estimates for the area in the town of West Hartford, CT

ISAT with Land Cover Coefficients

The Impervious Surface Analysis Tool (ISAT), an extension for ESRI’s ArcView and ArcGIS GIS software, was developed by the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center in collaboration with the University of Connecticut’s Nonpoint Education for Municipal Officials (NEMO) program for use by water resource managers and planners. The ISAT extension works in conjunction with the Spatial Analyst extension. ISAT requires three input ingredients: land cover data in an ESRI GRID format, analysis units (e.g., census tracts or watersheds) in an ArcGIS polygon shapefile format, and previously calculated impervious surface coefficients (one per land cover class). Impervious surface coefficients are derived by calculating the area of impervious surfaces (from calibration planimetric data) that fall within each land cover category. For each land cover class, its IS coefficient is the average percent of IS for this class among all calibration analysis units. To calculate the percent IS for each analysis unit, the ISAT overlays the polygon (analysis unit) data on land cover data and calculates the area of each land cover category within each polygon:

\[
IS_{AU} = \frac{\sum Area_i \cdot IS_i}{Total Area},
\]

where \(IS_{AU}\) is the percent IS for each analysis unit, \(Area_i\) is the area of the particular land cover category within this analysis unit, \(IS_i\) is the IS coefficient for this specific land cover category, and \(Total Area\) is the area of the analysis unit.

The set of IS coefficients, based on National Land Cover Data (NLCD) land cover, was generated at the University of Connecticut. The planimetric data, circa 2000, for the 42 towns in New York were used to calibrate a new set of coefficients, and updated planimetric data, circa 2002, for the ten Connecticut towns were used to test the coefficients at the Census tract level.

Land Cover and Population Density Regression

The land cover and population density-based regression model used in this study was created using the JMP Statistical Discover Software 5.0.1. As with ISAT, this model relied on NLCD 2001 land cover but included also census tract polygons with associated population values. A Fit Model application was used to perform Stepwise Regression analysis. Although the Stepwise feature produces estimates that are the same as those of other least squares analyses, it is capable of selecting among many models in order to find the most suitable one (SAS Institute, Inc., 2002). There were 22 independent variables selected: population density and the percentage of each of the 21 NLCD land cover classes within a given Census tract. Actual percent imperviousness was calculated from the union of the planimetric data layer with the tract...
boundaries, and the amount of each NLCD land cover class present in the study area was derived using the Tabulate Areas command from the ArcToolbox Spatial Analyst Tool for each tract. Westchester County, NY tracts (n = 221) were used as calibration data for regression analysis. Tracts from the ten Connecticut towns (n = 82) were used for testing and validation. Planimetric-based percent imperviousness, population density, and area values for each NLCD class were exported into the JMP input table. Values of the percent imperviousness per tract were selected as role variables, while the population density and area values were added as Model Effects. All land cover classes, except the Open Water and Wetland classes, were manually entered into the equation, as contributors to the regression model. For the remaining two groups (seven classes total), 0.25 was the significance threshold indicating that the corresponding regressor term was to be entered into the model (Tabachnick and Fidell, 2001).

A classical linear regression model was utilized, following the general equation:

\[ IS = b_1 + b_2 \cdot \text{PopDen} + \sum (b_i \cdot \%A_i) , \]

where \( b_1 \) is the IS-axis intercept, \( b_2 \) is the coefficient for population density expressed in persons per square mile, and \( b_i \) are those for the percentage of the NLCD category area within the tract.

RESULTS

The IS coefficients for the ISAT and population density-based methods were calibrated on the 221 New York tracts and validated on the 82 Connecticut tracts. In contrast with our previous study (Chabaeva et al., 2004), the calibration and validation sets came from different geographic regions. Table 1 contains the coefficients for the two methods. It should be noted that these coefficient sets are not stratified by population densities, as enabled by ISAT (high, medium, and low), but are the overall averages of imperviousness by land cover type. For the NLCD sub-pixel method, which does not use such coefficients, the IS values were extracted directly from the Connecticut portion of the NLCD IS dataset. For the ISAT model, four of the 20 land cover classes were not present in the New York calibration data or they contained no impervious features and so the corresponding IS coefficients were set to 0. For the population density-based model, 13 land cover classes were entered manually (it was deemed possible that those classes contained urban features smaller than the resolution of the NLCD 30 meter data), and another three classes (out of seven remaining) passed the 0.25 significance threshold.

To illustrate the comparative performance of the three methods, the town of Groton, CT was selected, as an area with diverse (among different tracts) land cover and population densities. Figure 5 provides the actual percent imperviousness for Groton based on the planimetric IS data. The maps in Figures 6, 7, and 8 show the predicted percent imperviousness for Groton (for the NLCD sub-pixel, ISAT, and population density-based methods, respectively). Each figure also contains the scatter plot (actual vs. predicted IS values), fitted trend line, \( R^2 \) value, and the overall RMSE value for the corresponding method (note: these plots, lines, and values are based on all 82 Connecticut tracts and not only those for Groton).

The RMSE values for the NLCD sub-pixel, ISAT, and population density-based models were 5.65, 5.48, and 4.56, respectively, indicating a better fit for the population density-based method. However, the \( R^2 \) value was found to be better for the NLCD sub-pixel method, 0.95 versus 0.93, for the other two methods. Looking at the scatter plots (Figures 6-8) it can be observed that the NLCD sub-pixel method over-estimates IS whereas the ISAT method underestimates IS resulting in slightly higher RMSE values. The population density-based method provides an overall better approximation of IS although there is more scatter resulting in a higher \( R^2 \) value.

To simplify the validation process, a model was created in ESRI’s ArcGIS Model Builder that allowed the application of NLCD and population density regression based coefficients to the 82 Census tracts in Connecticut. Land cover and an area unit shapefile (in our case Census tract 2000 file) were used as the source data for the model. Using the Tabulate Area command, the area of each landcover class for each unit was calculated and then the total unit area was calculated. The percent of each landcover was then calculated for each unit area. Although the model may require further coefficients calibration, it can be used to ease the application of the NLCD regression-based method.
Table 1. Regression coefficients for (a) estimating percent imperviousness from population density and percent coverage of NLCD land cover classes and (b) use with the Impervious Surface Analysis Tool (ISAT) and NLCD land cover

<table>
<thead>
<tr>
<th>Category</th>
<th>NLCD Class #</th>
<th>Regression NLCD</th>
<th>ISAT NLCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>11</td>
<td>0</td>
<td>0.4%</td>
</tr>
<tr>
<td>Developed Open Space</td>
<td>21</td>
<td>0.310582</td>
<td>13.9%</td>
</tr>
<tr>
<td>Low Intensity Developed</td>
<td>22</td>
<td>0.549918</td>
<td>30.3%</td>
</tr>
<tr>
<td>Medium Intensity Developed</td>
<td>23</td>
<td>0.570255</td>
<td>47.2%</td>
</tr>
<tr>
<td>High Intensity Developed</td>
<td>24</td>
<td>0.956886</td>
<td>62.8%</td>
</tr>
<tr>
<td>Bare Land</td>
<td>31</td>
<td>0.109363</td>
<td>34.6%</td>
</tr>
<tr>
<td>Unconsolidated Shore</td>
<td>32</td>
<td>24.780062</td>
<td>0.0%</td>
</tr>
<tr>
<td>Deciduous Forest</td>
<td>41</td>
<td>0.086670</td>
<td>3.7%</td>
</tr>
<tr>
<td>Evergreen Forest</td>
<td>42</td>
<td>0.262148</td>
<td>9.0%</td>
</tr>
<tr>
<td>Mixed Forest</td>
<td>43</td>
<td>0.189675</td>
<td>2.9%</td>
</tr>
<tr>
<td>Scrub/Shrub</td>
<td>52</td>
<td>0.547459</td>
<td>5.8%</td>
</tr>
<tr>
<td>Grassland</td>
<td>71</td>
<td>-0.314002</td>
<td>7.1%</td>
</tr>
<tr>
<td>Pasture/Hay</td>
<td>81</td>
<td>0.065397</td>
<td>8.7%</td>
</tr>
<tr>
<td>Cultivated</td>
<td>82</td>
<td>-1.192875</td>
<td>29.2%</td>
</tr>
<tr>
<td>Palustrine Forested Wetland</td>
<td>91</td>
<td>0.378831</td>
<td>1.1%</td>
</tr>
<tr>
<td>Palustrine Scrub/Shrub Wetland</td>
<td>92</td>
<td>2.739063</td>
<td>0.0%</td>
</tr>
<tr>
<td>Estuarine Scrub/Shrub Wetland</td>
<td>94</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Palustrine Emergent Wetland</td>
<td>96</td>
<td>-3.657089</td>
<td>1.3%</td>
</tr>
<tr>
<td>Estuarine Emergent Wetland</td>
<td>97</td>
<td>0</td>
<td>5.1%</td>
</tr>
<tr>
<td>Palustrine Aquatic Bed</td>
<td>98</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Intercept</td>
<td>N/A</td>
<td>-14.566984</td>
<td>N/A</td>
</tr>
<tr>
<td>Population Density</td>
<td>N/A</td>
<td>0.000085</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Figure 5. Actual imperviousness, Town of Groton, CT
Figure 6. NLCD Sub-pixel IS results summarized over 2000 census tracts, Town of Groton, CT. Scatterplot of predicted versus actual %IS for all CT study tracts.

Figure 7. ISAT IS results summarized over 2000 census tracts, Town of Groton, CT. Scatterplot of predicted versus actual %IS for all CT study tracts.

Figure 8. NLCD regression model IS results summarized over 2000 census tracts, Town of Groton, CT. Scatterplot of predicted versus actual %IS for all CT study tracts.
CONCLUSIONS

The study described in this paper marks the first occurrence of coefficient sets whose calibration and validation regions are from different states. As such, the study was an important test of the robustness of the two methods involved (the population density-based and ISAT models). The results were found to be no worse than those of the 2004 and 2006 study (Chabaeva et. al., 2004; Civco et al., 2006), where the same region was used for both purposes.

There are advantages and disadvantages to each of the impervious surface estimation methods examined. The higher accuracy achieved with the population and land cover-based regression model is especially appealing because of the wide availability of NLCD and population data. It is also fairly easy to implement within a GIS (the same is true for the ISAT), and can be adapted and recalibrated to different analysis units such as census blocks or watersheds.

The NLCD sub-pixel method, while seemingly less accurate, does offer the advantage of being spatially explicit – that is, it provides positionally-specific (at the pixel resolution) imperviousness estimates, rather than a homogenous (lumped) measure as do the other two methods.

Efforts continue to refine all of the techniques discussed in this paper, to extend their application geographically to other regions of the United States, and to implement them at a different size of analysis units.

ACKNOWLEDGEMENT

This paper is based, in part, upon work supported by the University of Connecticut’s Institute of Water Resources under the project Development of Regionally Calibrated Land Cover Impervious Surface Coefficients, and, in part, under the EPA Long Island Sound Study project Mapping and Monitoring Changes in Impervious Surfaces in the Long Island Sound Watershed. [CLEAR Publication Number 070215.1].

REFERENCES


ASPRS 2007 Annual Conference
Tampa, Florida    •    May 7-11, 2007


