

# DEVELOPMENT OF A POPULATION DENSITY AND LAND USE BASED REGRESSION MODEL TO CALCULATE THE AMOUNT OF IMPERVIOUSNESS

**Anna A. Chabaeva**, Graduate Research Assistant  
**Daniel L. Civco**, Director and Professor  
Center for Land use Education And Research  
Department of Natural Resources Management and Engineering  
College of Agriculture & Natural Resources  
University of Connecticut  
Box U-4087, Room 308, 1376 Storrs Road  
Storrs, CT 06269-4087  
[achabaev@canr.uconn.edu](mailto:achabaev@canr.uconn.edu)  
[dcivco@canr.uconn.edu](mailto:dcivco@canr.uconn.edu)

**Sandy Prisloe**, Geospatial Extension Specialist  
Center for Land use Education And Research  
UConn Cooperative Extension System  
1066 Saybrook Road, PO Box 70  
Haddam, CT 06438-0070  
[sprisloe@canr.uconn.edu](mailto:sprisloe@canr.uconn.edu)

## ABSTRACT

Nonpoint source pollution has been cited as the nation's number one water quality problem and urban runoff is the pollutant of greatest concern. It has been noted that the amount of urban runoff and its impacts on stream conditions and water quality are correlated to anthropogenic watershed imperviousness. While much research has focused on quantifying this relationship, little work has been done to develop methods to measure impervious surface area at the watershed-scale.

This paper provides details of research to develop an impervious surface estimation model for use with National Land Cover Data (NLCD) and Census Tract population density data. High-resolution planimetric data for buildings and transportation features served as the footprints for impervious surfaces from towns in Connecticut, Massachusetts, and New York and were used to calibrate the model. Census Population data most contemporaneous with the planimetric data were collected for the more than 100 tracts covering these towns. Percent imperviousness was calculated for each tract regressed against predictor variables of population density and NLCD class percentage per tract. A random sample of 80% of all the tracts was used to produce a linear regression model with an R-squared of 0.95, and RMSE less than 6%, with predicted imperviousness being within 1-2% for most tracts, when applied to independent validation data. A method has been developed to port this model to watersheds using interpolated population density and NLCD data. This has produced a regionally-applicable impervious surface estimation tool to be used by watershed managers.

Research results are being distributed through the national NEMO Network and are being shared with cooperators at NOAA Coastal Services Center to refine the Impervious Surface Analysis Tool (ISAT).

## INTRODUCTION

Nonpoint source pollution has been identified as one of the nation's greatest water quality problems, and runoff associated with urban development is the pollutant of major concern (Environmental Protection Agency, 1994). Impervious surface acts as a transportation artery for different types of water carried pollutants. Several researchers have noted the relationship between the amount of impervious surface and the water quality in watersheds (Schueler, 1994a; Arnold and Gibbons 1996). In order to reduce future water quality degradation, present amounts of imperviousness within communities must be measured, which can be easily done by town planners if they have high accuracy digital database containing

**ASPRS Annual Conference Proceedings**  
**May 2004 \* Denver, Colorado**  
*ASPRS – 70 years of service to the profession*

information on the built infrastructure (*i.e.*, roads, parking lots, sidewalks, driveways, buildings, etc.). In the absence of data that permit the direct measurement of imperviousness, it can be modeled using surrogates such as land use or population, data that are often more widely available. The research reported here used high resolution planimetric data to calibrate and validate a model to predict the degree of impervious cover using land cover and population density.

## IMPERVIOUS SURFACE MEASUREMENT

There are a number of different techniques that can be used to measure or estimate the area of impervious surfaces. The most time consuming and costly, but on the other hand often the most accurate, is manual extraction of impervious surface features from remote sensing imagery through *heads up* digitizing. Point sampling (Bird *et al.*, 2001) can be used as an alternative to digitizing, but personal experience has found this also to be very time consuming yet less accurate. Remote sensing pattern recognition approaches such as sub-pixel classification (Civco *et al.*, 2002, Justice *et al.*, 2003), artificial neural networks (Mokken, 1996, Flanagan and Civco 2001), Classification and Regression Trees (CART) (Yang, 2003, Herold, 2003), and index based methods and models such as the Normalized Vegetation Difference Index (NDVI) (Carlson and Ripley, 1997, Hebble *et al.*, 2001, and Xie *et al.* 2003) and Vegetation-Impervious Surface-Soil (VIS) (Hung and Ridd, 2002) have been used for measuring imperviousness, but require both moderate to high resolution remote sensing data as well as the expertise to process and analyze them. These data and analytical capabilities are often beyond the reach of many planners and decision makers at the local level.

There have been several efforts to determine the relationship between the percent of impervious surface and different urban development parameters such as road density, population density, land use type and size of development units (Woods *et al.*, 1999; Sleavin, 1999; Prisløe *et al.*, 2003). Previous research at the University of Connecticut investigated the relationship between impervious surface area and a 1995 Connecticut Landsat-derived land use land cover dataset. It also was noted, although not thoroughly researched at the time, that there appeared to be a positive correlation between population density within a land use land cover class and the amount of impervious surface.

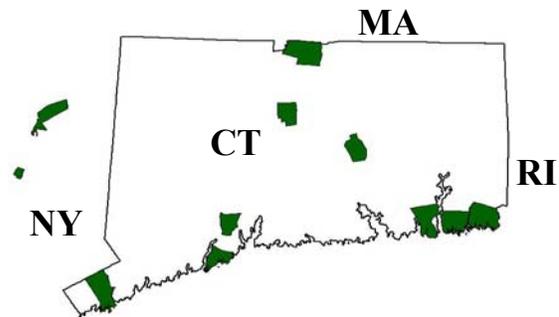
## RESEARCH OBJECTIVES

The overarching objective of this research was to develop a regionally-calibrated set of impervious surface coefficients for land use and land cover categories for widely available datasets, such as the National Land Cover Data (NLCD) or NOAA's Coastal Change Analysis Program (C-CAP), thereby allowing land use planners and water resource managers to calculate quickly existing levels of impervious cover in a community or watershed. The specific objectives of the research reported here were to (a) determine the quantitative relationship between percent imperviousness with land use and population, (b) develop, assess, and refine a regression-based model to predict percent imperviousness at the census tract level, and (c) extrapolate this model to the watershed unit and extend it to the entire state of Connecticut.

## DATA AND PREPROCESSING

### Study Area

The calibration and validation data were chosen on the basis of availability of planimetric data for municipalities in Connecticut, Massachusetts and New York: Groton, Marlborough, Milford, Stonington, Stamford, Suffield, Waterford, West Hartford and Woodbridge in Connecticut, Amherst in Massachusetts, and North Castle and Mount Vernon in New York (Figure 1). The municipalities vary from rural to highly urbanized and represent a good cross section of development patterns within the study area.



**Figure 1.** Location of study area towns in Connecticut and New York. Not depicted is Amherst, located in northern Massachusetts.

## Data

Data required for the development of the model included: planimetric impervious feature data for calibration and validation, land use and land cover data, and population (density) data. Further, initial analyses were bounded by US Census Bureau Tracts, and subsequently were extended to watershed units.



**Figure 2.** Sample planimetric data.

Census tract boundaries, which include many positional inaccuracies, were edited as described below.

The digital *planimetric data* generally were circa 1995; however some municipal datasets were a year or two older and several were more recent. These data represent the footprint of impervious surface features within the study area (Figure 2).

There are two major groups of features that can be labeled as impervious: rooftops and the various components of the transportation system. Structures such as buildings, pools and patios fall within the rooftop component of imperviousness. Roads, sidewalks, driveways and parking lots are included in the transportation system. Features such as rock outcrops and barren lands that are not anthropogenic, but still can be considered as impervious surface, were not included in the analysis.

National Land Cover Data (NLCD) were obtained from the United States Geologic Survey (USGS), and were chosen because of their nationwide availability, thereby enabling the geographic extension of the model under development to parts of the country other than the Northeast. Once processed, this data set had a resolution of 100 by 100 feet. NLCD contains 22 land cover categories, 17 of which were present in the area of study.

Digital 1990 *Census tract boundaries* for Connecticut, Massachusetts and New York and *Population data* for the 1990 census tracts in each state were obtained from the U.S. Census Bureau web site. The municipal boundary data for Connecticut were downloaded from the Map and Geographic Information Center (MAGIC) at University of Connecticut. *Watershed data* were acquired from the Connecticut Department of Environmental Protection.

## Preprocessing Procedures

Where necessary, the boundary for each municipality was edited to exclude significant areas of water. This applied especially to those municipalities bordering Long Island Sound and to those that had a large river within or bounding the municipality.

Considerable editing was done to the Census tracts. Tract boundaries were adjusted to match road centerlines, municipal boundaries and waterbody shorelines as depicted on the municipal planimetric data and the edited municipal boundaries datasets. These edits were necessary to insure that tracts could be overlaid accurately with planimetric and other digital datasets.

The NLCD data were downloaded in a *grid* format from the USGS, and were subsequently converted into an ESRI shapefile with ArcView version 3.2 and the Spatial Analyst Extension using a simple Avenue script. (Note: the script used a “no weed” option that preserved grid edges.) NLCD shapefiles were then clipped to the boundary of each municipality. Each polygon contained information on the area of NLCD classes present.

All planimetric feature 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.). All the impervious feature shapefiles were clipped with the appropriate town boundary.

Population density in people per square mile was calculated from the Census Bureau 1990 population tables using the edited area of each tract within the study municipalities.

A watershed dataset was created to eliminate large out-of-state watersheds that had only a small area within Connecticut. Only watersheds, which had their center within the boundary of the state, were selected from the Connecticut watershed dataset.

## METHODOLOGY

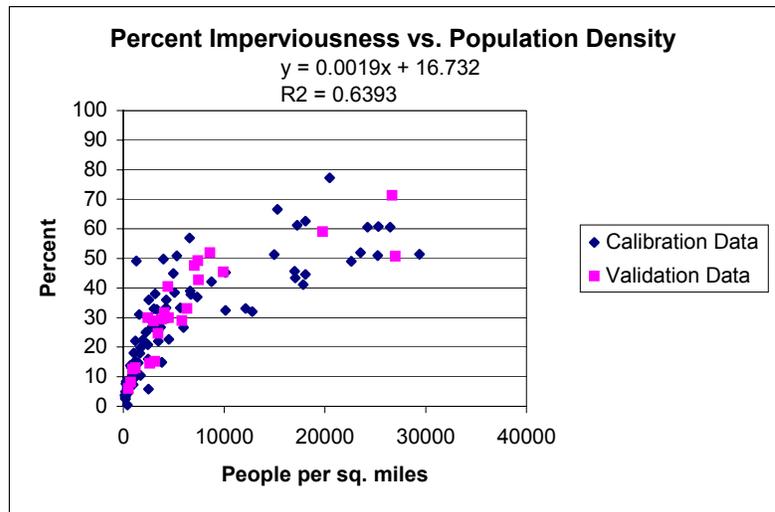
Land use specific impervious surface coefficients were derived using a method developed at the University of Connecticut. NLCD shapefiles for each tract were combined with the planimetric data using ArcView's Geoprocessing *Union* operation. This resulted in a new shapefile where each polygon had two values: NLCD class and a binary value to indicate if the polygon was impervious or not. For each tract, two frequency tables were created using XTOOLS, an ArcView Extension developed by the Oregon Department of Forestry<sup>1</sup>: (1) the sum of the area of each land cover class, and (2) the sum of the impervious surface area for each land cover class. The tables were exported to Excel and impervious surface coefficients, which are the percent area of each land cover class that is covered with impervious features, were calculated using the following formula:

$$LULC_{coefficient} = \frac{LULC_{ISarea}}{LULC_{Area}} \cdot 100\%$$

where  $LULC_{ISarea}$  is the total area of impervious surface for the NLCD class, and  $LULC_{Area}$  is the total area for the same NLCD class.

This procedure was repeated for each of the 108 tracts within the study area, resulting in a database containing the total area of the tract, the percent impervious surface area of the tract, the area of each land cover category within each tract and its calculated impervious surface coefficient. The Census tract 1990 population density data were added to the database. The percent impervious surface for each tract was plotted with the population density data (Figure 3). The strong positive relationship between percent IS cover and population density is well illustrated, especially for lower densities and levels of imperviousness. The inclusion of a land cover variable will serve to strengthen this relationship.

The JMP Statistical Discover Software<sup>2</sup> was used to create a regression model for calculating the amount of imperviousness. There were 17 variables used for the *Fit Model* application of JMP: population density and the percentage of each land use land cover class per tract. Due to the non-linear nature of the data, land use percentages were transformed with the square root of this value and population density was transformed to its Base<sub>10</sub> logarithm. Eighty percent of the tracts were randomly selected from the sample, upon which multivariate statistical analysis was applied and a regression equation derived. The remaining tracts were used for the testing and validation.



**Figure 3.** Population Density vs. Percent Imperviousness

## RESULTS AND DISCUSSION

The regression analysis resulted in a set of coefficients that predicts, at the tract level, percent imperviousness as a function of population density and percent coverage of the NLCD land use types (Table 1). The Fit Model automatically omitted from the regression formula several of the NLCD classes

<sup>1</sup> [http://www.odf.state.or.us/DIVISIONS/management/state\\_forests/XTools.asp](http://www.odf.state.or.us/DIVISIONS/management/state_forests/XTools.asp)

<sup>2</sup> <http://www.jmp.com/>

due to their statistical insignificance. Of the 17 NLCD land cover classes found in the study areas, 10 were observed to be significant contributors. They are listed in Table 1.

Variable Description	Variable	Coefficient
Intercept	$b_1$	1.61051845
Population Density	PopDen	0.93584385
High Intensity Residential	$A_{22}$	0.20444150
Commercial/Industrial/Transportation	$A_{23}$	0.28746228
Bare Rock/Sand/Clay	$A_{31}$	-0.20257370
Deciduous Forest	$A_{41}$	-0.26362550
Evergreen Forest	$A_{42}$	-0.20714210
Shrubland	$A_{51}$	1.18131149
Orchards/Vineyards/Other	$A_{61}$	1.87563299
Pasture/Hay	$A_{81}$	0.17330553
Row Crops	$A_{82}$	-0.47697190
Emergent Herbaceous Wetlands	$A_{92}$	-0.22767610

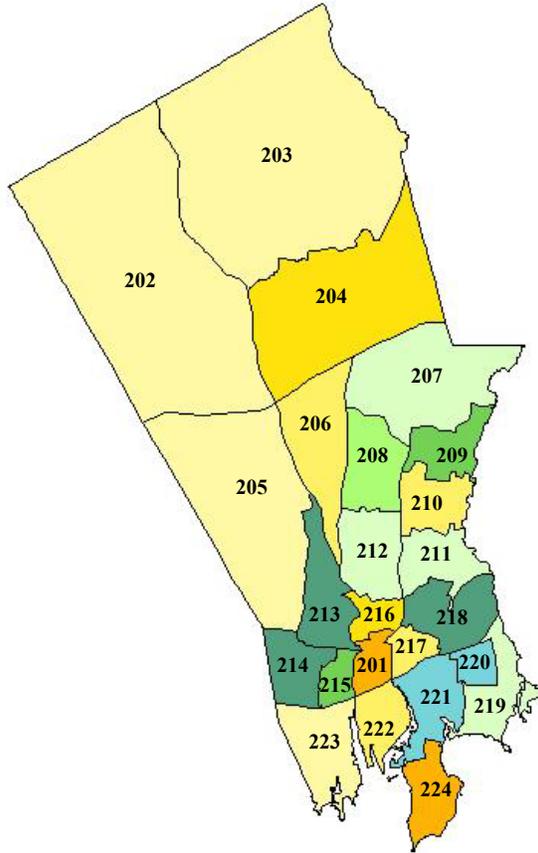
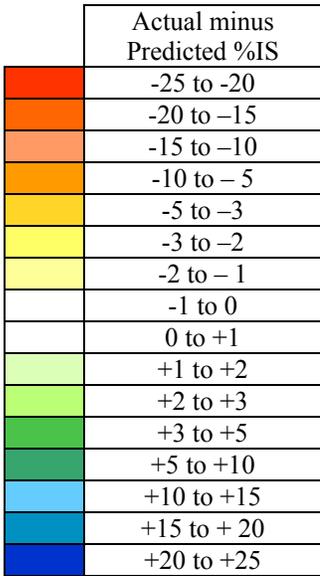
**Table 1.** Regression model coefficients.

These coefficients can be used in a regression model that was created based on the classical linear regression and can be described by the following equation:

$$\text{Percent of Imperviousness} = b_1 + b_2 \cdot \lg \text{PopDen} + b_3 \cdot \sqrt{\%A_{22}} + b_4 \cdot \sqrt{\%A_{23}} + \dots + b_{19} \cdot \sqrt{\%A_{92}}$$

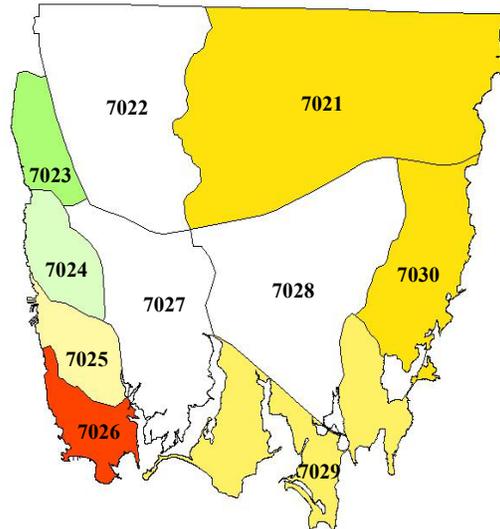
where  $b_1, b_2, b_3, \dots, b_{19}$  are the regression coefficients, *PopDen* is the Population density and  $\%A_{22}, \%A_{23}, \%A_{31}, \dots, \%A_{92}$  are the percent of the NLCD category area within the tract. This regression equation was tested on the remaining 20 percent of the tracts. The comparison of the regression model results with the actual impervious surface area measured from planimetric data for each tract yielded an  $R^2$  of 0.95 and an RMSE of less than 6%, when applied to validation data. The model then was applied to all the tracts and the actual and predicted values of imperviousness were compared. This resulted in an overall RMSE of 4% (Figure 4). For tract number 7026 in Groton, CT imperviousness was underestimated by 22 percent. This is likely due to the presence of industrial facilities (General Dynamics and Pfizer) and the University of Connecticut at Avery Point, a non-residential campus.

### Stamford, CT



Tract	PI ac	PI pr	Diff
201	77	70	-7
202	7	6	-1
203	11	10	-1
204	13	10	-3
205	15	14	-1
206	28	26	-2
207	20	22	2
208	27	30	3
209	38	43	5
210	36	34	-2
211	48	50	2
212	33	35	2
213	32	38	6
214	52	60	8
215	59	63	4
216	63	60	-3
217	67	65	-2
218	45	52	7
219	33	35	2
220	46	57	11
221	49	60	11
222	51	49	-2
223	45	44	-1
224	29	22	-7

### Groton, CT

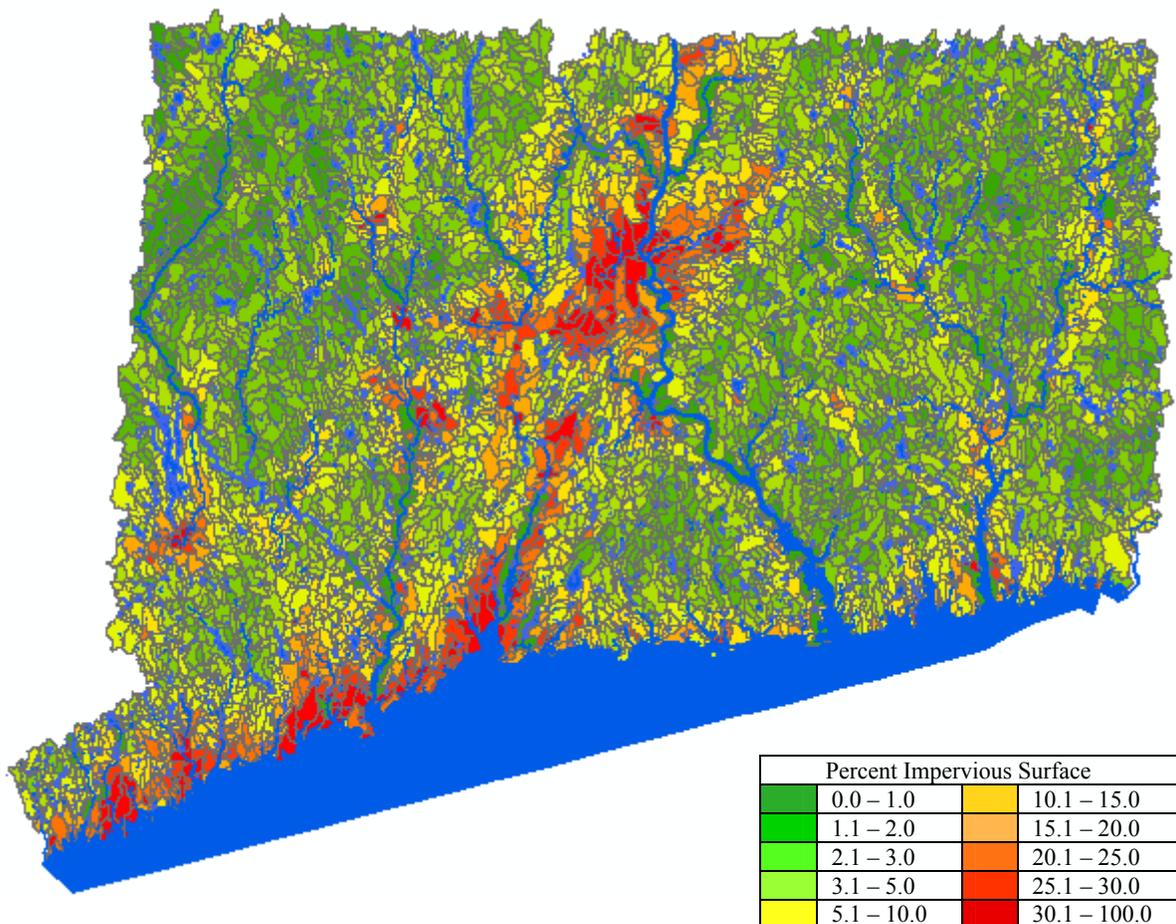


Tract	PI ac	PI pr	Diff
7021	6	3	-3
7022	11	11	0
7023	23	26	3
7024	29	31	2
7025	39	38	-1
7026	49	27	-22
7027	22	22	0
7028	9	10	1
7029	10	8	-2
7030	15	12	-3

Figure 4. Difference between actual (PI ac) and predicted (PI pr) percent imperviousness at the tract level for the municipalities of Stamford and Groton, CT.

To test further the impervious prediction model, a procedure was developed to apply it at the watershed-level, a common and meaningful resource management unit. For Connecticut, there are 6,711 watersheds ranging in size from 0.4 to 13481 acres, with an average size of 474 acres. To apply the regression model, we required the area of each NLCD class within each watershed and a population density value. Calculating NLCD class areas was done in the same manner as was done for the Census tracts, as previously described. However, it was necessary to develop a methodology to approximate population density for each watershed. Census block and population data for 2000 were used to calculate the population density at the block level. The Census block boundaries then were converted to a grid format having the same origin, grid cell size and extent as the NLCD data and the population density values were saved as grid cell values. The ArcView Spatial Analyst Extension was used with the gridded Census block population density data summarized over the zones defined by the areal extent of each watershed. Watersheds that fell completely within study municipalities for which we had planimetric impervious feature data were selected and the actual percent impervious surface area within these watersheds was calculated.

The Census tract based regression model was applied to all the watersheds and the percent impervious surface area for each watershed was calculated and plotted (Figure 5).



**Figure 5.** Predicted percent impervious surface for Connecticut watersheds.

The results of the regression model were compared to the actual percent impervious surface area for the 236 watersheds for which we had complete coverage with planimetric data (Figure 6). The RMSE was less than 4% and many watersheds were within 1% to 2% of the actual value.

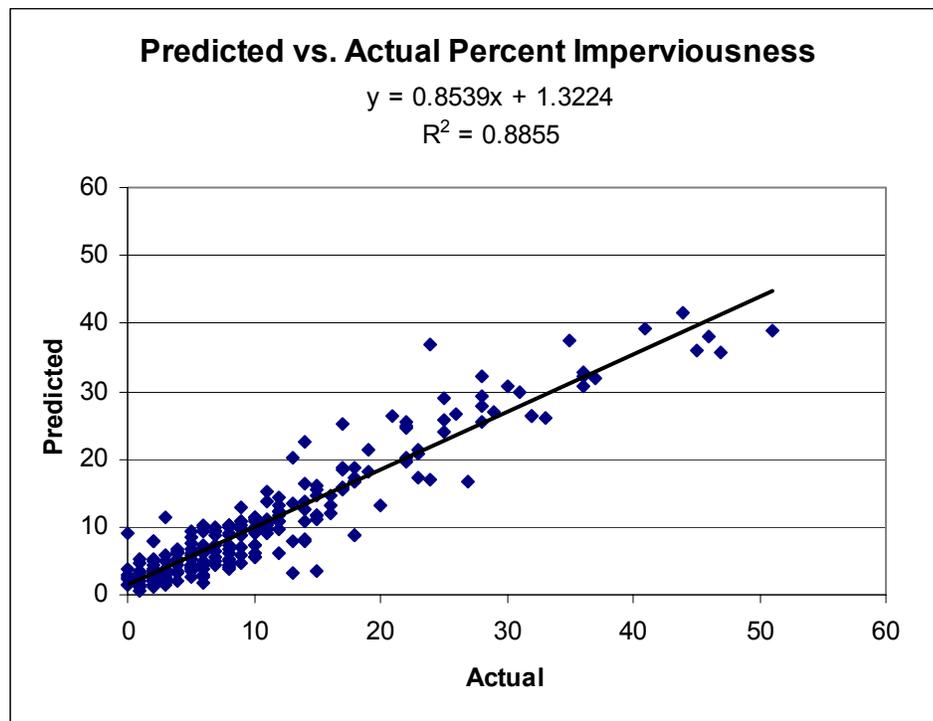


Figure 6. Predicted vs. actual percent imperviousness for Connecticut watersheds.

## CONCLUSION

The regression model developed through this research does an accurate job of predicting the percent area of a Census tract or a watershed that is covered with impervious surfaces. It uses readily available National Land Cover Data (NLCD) and US Census Bureau population data (converted to population density) as inputs. The model, developed on the basis of Census tract geography, was tested for watersheds and the results were similar indicating that it also may be applicable over other geographic areas.

The model's output will be of value to land use planners and water resource managers in Connecticut and in others areas with similar landscape patterns. It provides an easy to use and accurate method to estimate overall imperviousness for areas where detailed planimetric data are unavailable. Such information is critically important to the design and implementation of land use plans and strategies aimed at preventing water quality impairment from urban non-point source pollution.

It is important to recognize that the model described here has been tested and validated only in Connecticut. It is possible, and may very well be likely, that variables generated by the regression analysis are not generally extensible to other geographic areas where development and land cover patterns are different. Future research needs to be conducted in this area. It is our hope that researchers with access to similar data in other areas of the country will use the methodology described in this paper to develop their own regionally calibrated set of regression model variable values. It is conceivable that over time libraries of such values could be created to aid planners and managers throughout the country.

The authors did not evaluate the regression model's sensitivity to geographic scale. There may be scale-dependencies that escaped notice given the relatively uniform size of Census tracts and watersheds that were used in the study. Additional research is warranted to assess these possible influences.

## ACKNOWLEDGMENTS

The authors thank the local town planners for providing planimetric data used in this study. This material is based 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". [CLEAR Publication Number 040227.1]

## REFERENCES

- Arnold, C.L., and Gibbons, C.J. (1996). Impervious Surface Coverage: The Emergence of a Key Environmental Indicator. *Journal of the American Planning Association*, 62 (2): 243-258.
- Bird, S.L., and Exum, L.R. (2001). Generating High Quality Impervious Cover Data. *Quality Assurance*, 8: 91-103.
- Carlson, T.N., and Ripley, D.A. (1997). On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index. *Remote Sensing Environment*, 62: 241-252.
- Civco, D.L., Hurd, J.D., Wilson, E.H., Arnold, C.L., and Prisloe, M.P. (2002). Quantifying and Describing Urbanizing Landscapes in the Northeast United States. *Photogrammetric Engineering and Remote Sensing*, 68(10): 1083-1090.
- Environmental Protection Agency. (1994). The Quality of Our Nation's Water: 1992. United States Environmental Protection Agency #EPA-5-94-002. Washington, D.C.: USEPA
- Flanagan, M.C., and Civco, D.L. (2001). Subpixel Impervious Surface Mapping. ASPRS 2001 Annual Convention, St. Louis, MO, April 23-27.
- Justice D., Rubin, F. (2003). Developing Impervious Surface Estimates for Coastal New Hampshire. A Final Report to the New Hampshire Estuaries Project.
- Herold, N.D, Koeln, G., and Cunnigham, D. (2003). Mapping Impervious Surfaces and Forest Canopy using Classification and Regression Tree (Cart) Analysis. ASPRS 2003 Annual Conference Proceedings.
- Hebble, E.E, Carlson, T.N., and Daniel, K. (2001). Impervious Surface Area and Residential Housing Density: A Satellite Perspective. *Geocarto International*, 16(1).
- Hung, M-C., and Ridd, M.K. (2003). A Subpixel Classifier for Urban Land-Cover Mapping Based on a Maximum-Likelihood Approach and Expert System Rules. *Photogrammetric Engineering and Remote Sensing*, 68(11): 1173-1180.
- Mokken, R.J. (1996). Remote Sensing, Statistics and Artificial Neural Networks – some R and D Perspectives. In Proceedings of the International Symposium on Spectral Sensing Research 1995. CCSOM Reprint 95-140.
- Prisloe, M., Wilson, E.H., and Arnold, C. (2003). Refinement of Population-Calibrated Land-Cover-Specific Impervious Surface Coefficients. Final Report. NEMO FY '02 Work Plan DEP Project 01-08 Task #6. University of Connecticut Middlesex County Extension Center.
- Schueler, T.R. (1994a). The importance of Imperviousness. *Watershed Protection Techniques*, 1(3): 100-11.
- Sleavin, W.J. (1999). Measuring Impervious Surfaces in Connecticut Using Planimetric GIS Data. M.S. Thesis, Department of Natural Resources Management and Engineering, The University of Connecticut, Storrs, CT.
- Wood, S. et al. (1999). Assessment of Current and Future GVS&DD Area Watershed and Catchment Conditions. Greater Vancouver Sewerage and Drainage District Liquid Waste Management Plan.
- Xie, H., Granillo, J.A., and Keller, G.R. (2003). Classification and Application of Impervious and Pervious Surface using Integrated Remote Sensing and GIS Technologies. ASPRS 2003 Annual Conference Proceedings.
- Yang, L., Huang, C., Homer, C.G., Wylie, B.K., and Coan, M.J. (2003). An Approach for Mapping Large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and High Spatial Resolution Imagery. *Journal of Remote Sensing*, 29(2): 230-240.