Journal of Environmental Science and Engineering A 4 (2015) 311-328
doi:10.17265/2162-5298/2015.06.006

Monitoring Local Impervious Surface Trends Using Indices Derived from Property Tax and Several Landsat Datasets

Marek Topolski
Department of Natural Resources, Fisheries Service, Maryland 21401, USA

Abstract: Human economic and social needs can be in conflict with ecosystem needs. Land development increases impervious surfaces causing significant negative impacts to aquatic ecosystems. Many impervious surface estimates are derived from remote sensing data, developed by using different methods and often out of date. Remote sensing data is often at scales applicable to regional management, but not local planning decisions. To date, no standardized annual dataset of percent impervious surface exists for use at both local and watershed scales. Effective communication between natural resource managers and local planners has been lacking. One solution is to monitor percent impervious surface with a relative index rather than direct measure. A relative index model can use a currency, like foundation square feet per hectare, which is useful for all decision makers. One data source for developing a relative index of impervious surface is property tax data. These data document annual land development at local scale. Here, the author presents the use of Maryland property tax data to index land development and percent impervious surface.

Key words: Geographically weighted regression, impervious surface, Landsat, MdProperty View, relative index, spatial analysis, tax records.

1. Introduction

Land development in the form of impervious surfaces (IS) such as roads, parking lots and buildings increases concentrations of pollutants like nitrogen, phosphorus, heavy metals and suspended solids in storm water [1, 2]. While, reduce infiltration of water into soil and rapidly delivering it to aquatic systems resulting in significant habitat alteration [3-6]. Percent IS, therefore, provides an index of these multiple stressors that are indicative of urbanization. Direct measure of IS using aerial photographs is the most accurate method to determine %IS, but this method is time intensive for relatively small catchments making it impractical and costly for county wide planning [7]. Satellite based remote sensing imagery, such as that collected by Landsat Thematic Mapper (TM), is a lower cost alternative to estimate %IS [8]. While commonly used, there are drawbacks. Development of %IS estimates from remote sensing data requires significant expertise [9-11] to correct for external factors that increase classification error including land cover heterogeneity, patch size of land cover, sensor resolution, spectral similarity among different feature types and classification methodology [12-14]. The complexity associated with interpreting satellite based data prohibits its use as an annual estimation of %IS changes by local land use decision makers.

Indexing %IS with another more readily available landscape measure is one option. An index is a derived variable based on empirical measures of other variables [15] from which a functional relationship can be defined [16]. Index reference points can be established for habitat thresholds such as critical %IS values for brook trout (*Salvelinus fontinalis*) [17], southern two-lined salamanders (*Eurycea cirrigera*) and northern dusky salamanders (*Desmognathus fuscus*) [18] and minimum dissolved oxygen levels for

Corresponding author: Marek Topolski, natural resources biologist, research field: fish habitat. E-mail: marek.topolski@maryland.gov.

fish and shellfish [16]. In general, increased %IS is associated with reduced aquatic habitat quality and changes to aquatic community composition [4, 7, 9, 10, 19], suggesting that a %IS index would be a useful tool for land use planning and management [4, 8].

Accuracy of %IS estimates varies among land uses and aggregation of similar land cover types. Land Use and Land Cover (LULC) classifications are often based on the four tier Anderson classification system [20]. Land types are classified into categories such as residential, manufacturing, transportation, resource extraction and undeveloped. Categories are subdivided as the Anderson level increases from I to IV. Effective use of the Anderson classification system requires a LULC classification accuracy ≥ 85% [20].

Overall classification accuracy of the 1992 National Land Cover Database (NLCD) can varry between 46% and 60% at Anderson level II and between 70% and 83% at Anderson level I [21]. For areas smaller than 10% IS, it predicted that %IS was consistently underestimated when using Landsat TM imagery compared to aerial photographs [5, 17]. Percent IS estimation error may result from land cover heterogeneity and patch size induced classification error [13, 22]. Hu and Weng [12] and Ji and Jensen [6] significantly improved IS interpretation accuracy of remote sensing imagery by using subpixel classifiers to detect subpixel level differences in spectral signature which enable the detection of IS in wooded and low-density residential areas. However, the subpixel classifier analyses excluded water, wetland, forest and grassland LULCs. Also, subpixel spectral analyses are unable to accurately interpret shaded areas and different but spectrally similar surfaces (dry dirt versus bright IS). Thus, make them most useful for urban areas [12]. Bauer et al. [8] used high resolution aerial photography and orthophoto quarter quads of Minnesota's twin cities—Metropolitan Area (TCMA) to improve Landsat TM %IS estimates to 3.5% of actual %IS. Despite improved accuracy, Bauer et al. [8] cautioned that time and imagery costs make the approach best suited for city scale analyses rather than larger areas such as the entire 7,700 km^2 TMCA.

Alternate data sources have been used to estimate land use change and %IS within an area. For example, Reilly et al. [11] incorporated measures of employment and home construction into a predictive %IS model. Bird et al. [23] and Exum et al. [5] integrated population density and major road IS with remotely sensed land cover estimates. Moglen and Beighley [24] used Maryland property tax records and maps to determine land use change and model peak discharge into streams within urbanizing watersheds. Wu, Silvanhyphen, and Wang [25] used building count and building shape from tax data to estimate %IS within a 250 square mile section of Austin, Texas. Stone [26] modeled %IS using parcel attributes (lot size, lot frontage, front yard setback and residential capacity) and street network design (street width and street intersection density).

Property-based models are promising, considering that impervious surface area increases with building and infrastructure construction [11]. Tax records document structure presence and size for each land parcel. Tax map data are standardized, annually updated and readily accessible. Considering that tax map data tracks the presence and absence of structures on land parcels, these data should be useful for developing an index of %IS. Multiple remotely sensed %IS datasets exist for Maryland. It is likely that each %IS dataset will result in a unique tax map index [16]. This study evaluated the feasibility of developing a tax map based index for %IS. These indices would provide a range of %IS status and trends that could inform local land use and resource management decisions.

2. Materials

2.1 Data Development

Geo-processing was done using ESRI ArcMap 10.1 and the Spatial Analyst extension [27]. All files were projected in NAD 1983 State Plane Maryland FIPS 1900 (Meters). Maryland tax data is available by county as point feature shapefiles from the Maryland Department of Planning (MDP) as part of the MdProperty View dataset [28]. All tax datasets used in the analysis included attributes for x, y coordinates, the year a primary structure was built and the primary structure's foundation footprint in square feet (Table 1). Each year's tax data was compiled into a single statewide file prior to geo-processing. Tax records without x, y coordinates and year built data were removed. Tax data has been updated annually since 1996. However, data from 1996 to 1998 did not include the century in which a structure was built (1700s, 1800s or 1900s) and were not used in the analysis. Tax data for years prior to 1999 were derived from the 1999 dataset. Tax data were queried for records between 1700 and the years 1990, 1992, 1996, 2000, 2001, 2005 and 2006 which corresponded with available %IS datasets for Maryland.

Maryland Department of Natural Resources (MDNR) created the shapefile SWSHED [29], herein 12 digit to delineate third-order watersheds (Fig. 1) based on the Strahler designation method [30] and adjusted based on U.S. Geological Survey (USGS) 7.5 minute quadrangle map sheets. Watersheds that exceeded 6,070 ha were split into subwatersheds when possible. Maryland 12 digit watersheds are comparable to USGS 12 digit hydrologic unit codes (HUCs).

Table 1 Number of tax records by Maryland county and the number which had x, y coordinates from the 2006 MdProperty View data. The number of records used based on structure year built or foundation footprint data is indicated. Percent of tax records are in ().

County	Total number of tax records	Records with x, y coordinates,		
			with structure (year built),	
				with structure size (ft^2)
Allegheny	41,013	39,487 (96)	27,502 (67)	27,481 (67)
Anne Arundel	202,254	200,248 (99)	169,087 (84)	168,980 (84)
Baltimore city	235,081	234,579 (100)	167,994 (71)	166,591 (71)
Baltimore	292,309	287,267 (98)	234,217 (80)	234,028 (80)
Calvert	41,066	40,596 (99)	30,889 (75)	30,542 (74)
Caroline	16,218	16,153 (100)	11,857 (73)	11,852 (73)
Carroll	65,206	64,395 (99)	55,136 (85)	54,730 (84)
Cecil	45,465	44,922 (99)	32,780 (72)	32,418 (71)
Charles	58,146	57,444 (99)	46,636 (80)	46,633 (80)
Dorchester	22,037	21,765 (99)	14,265 (65)	14,121 (64)
Frederick	89,504	87,264 (97)	76,215 (85)	72,584 (81)
Garret	28,352	27,941 (99)	15,977 (56)	15,929 (56)
Harford	93,137	91,833 (99)	79,662 (86)	78,945 (85)
Howard	95,516	94,650 (99)	84,236 (88)	84,104 (88)
Kent	13,152	12,971 (99)	8,786 (67)	8,767 (67)
Montgomery	325,015	323,341 (99)	293,172 (90)	292,240 (90)
Prince George	279,807	276,493 (99)	234,692 (84)	230,676 (82)
Queen Anne	24,456	24,033 (98)	18,444 (75)	18,239 (75)
Somerset	16,944	16,863 (100)	9,319 (55)	9,161 (54)
St. Mary's	44,775	43,765 (98)	32,980 (74)	32,977 (74)
Talbot	20,302	20,165 (99)	16,268 (80)	16,260 (80)
Washington	57,866	57,117 (99)	46,575 (80)	46,338 (80)
Wicomico	45,451	45,162 (99)	33,695 (74)	33,500 (74)
Worcester	64,215	63,049 (98)	50,832 (79)	50,675 (79)
Statewide	2,217,287	2,191,503 (99)	1,791,216 (81)	1,777,771 (80)

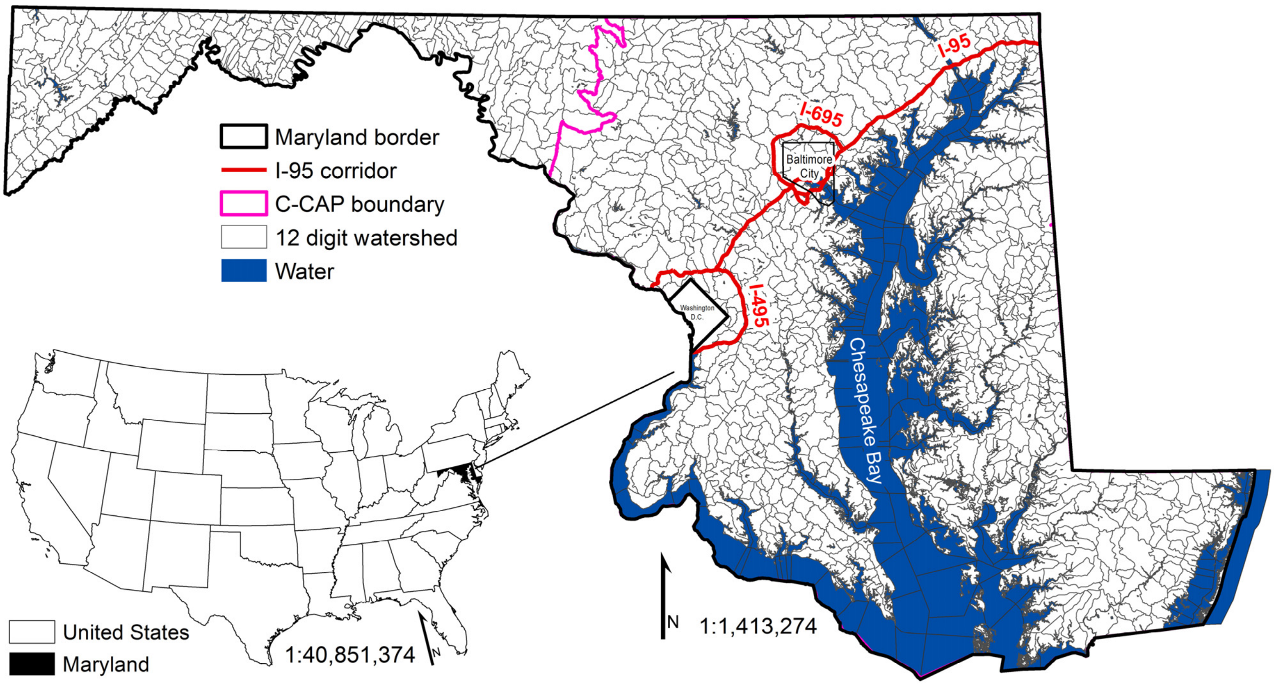

Fig. 1 Map of 12 digit watersheds including ponds, lakes, rivers and Chesapeake Bay in Maryland, USA.

Maryland tax data excludes Washington D.C. Portions of 12 digit watersheds within Washington D.C. were removed. Watersheds split by Washington D.C. boundary were recombined. A total of 1,120 watersheds remained. Since IS is the manifestation of terrestrial development patterns, all bodies of open water including estuaries, rivers, reservoirs, lakes and ponds (≥ 1.2 ha) were removed from each 12 digit watershed and all subsequent calculations. Removal of open water ensured that land use impacts in watersheds having extensive shorelines were not underestimated. Tax records were spatially joined with the 12 digit watershed containing it. Records having x, y coordinates outside of Maryland were joined to the closest 12 digit watershed. Total structure foundation square feet per hectare (ft^2/h) was calculated for each 12 digit watershed.

Four %IS and LULC datasets were compared with the tax data: NLCD [31-33], Coastal Change Analysis Program (C-CAP) [34], Mid-Atlantic Regional Earth Science Applications Center (RESAC) [35, 36] and Towson University (TU) [37]. Each %IS dataset was derived from 30 m × 30 m resolution Landsat TM imagery. Each dataset used different methodologies to produce %IS or LULC estimates, but each was validated with aerial orthophotographs. Percent IS estimates (Table 2) were on a continuous scale from 0% to 100%. However, LULC estimates were categorical requiring conversion to comparable %IS values for developed land from NLCD [38]. All other LULC classifications were from MDP [39].

The NLCD was produced by the multi-resolution land characterization—consortium's regional land cover characterization project [40]. Maryland NLCD data is a subset of the federal region-III project area. Separate leaves-on and leaves-off mosaic images were created from Landsat TM bands 3, 4, 5 and 7. A %IS estimate between zero and 100 was determined for each raster pixel through an iterative unsupervised and supervised classification process. Percent IS estimates were validated and corrected with other information sources such as: aerial photographs, census of agriculture, digital terrain elevation data, LULC and national wetlands inventory. NLCD 1992

Table 2 LULC categories for C-CAP and MDP datasets.

C-CAP code	C-CAP label	Mean C-CAP IS coefficient	MDP code	MDP label	MDP IS coefficient
1	Unclassified	-	11	Low density Residential	0.14
2	Developed, high intensity	0.90	12	Medium density Residential	0.28
3	Developed, medium Intensity	0.65	13	High density Residential	0.41
4	Developed, low intensity	0.35	14	Commercial	0.72
5	Developed, open space	0.10	15	Industrial	0.53
6	Cultivated crops	0.02	16	Institutional	0.34
7	Pasture/hay	0.02	17	Extractive	0.02
8	Grassland/herbaceous	0.01	18	Open urban land	0.09
9	Deciduous forest	0.01	21	Cropland	0.02
10	Evergreen forest	0.01	22	Pasture	0.02
11	Mixed forest	0.01	23	Orchards	0.02
12	Scrub/shrub	0.01	24	Feeding operations	0.02
13	Palustrine forested wetland	0.00	242	Agricultural building	0.02
14	Palustrine scrub/shrub wetland	0.00	25	Crops	0.02
15	Palustrine emergent wetland (persistent)	0.00	41-44	Forest/brush	0.00
16	Estuarine forested wetland	0.00	50	Water	0.02
17	Estuarine scrub/shrub wetland	0.00	60	Wetlands	0.00
18	Estuarine emergent wetland	0.00	71	Beaches	0.00
19	Unconsolidated shore	0.00	72	Bare rock	0.09
20	Barren land	0.09	73	Bare ground	0.09
21	Open water	0.00	191, 192	Rural residential	0.04
22	Palustrine aquatic bed	0.00	-	Highway	0.95
23	Estuarine aquatic bed	0.00	-	-	-
24	Tundra	0.01	-	-	-
25	Perennial ice/snow	0.01	-	-	-

* C-CAP coefficients are from NOAA Coastal Services Center (n. d.) and MDP coefficients are from the Center for Watershed Protection (2005, p. 64). IS from highways and roads were not included in this analysis.

data were classified as one of 23 categories developed from the C-CAP classification protocol and federal geographic data committee standards [14]. NLCD 1992 LULC classifications were reclassified to the corresponding NLCD 2001 mean %IS coefficient [41] as described by Bird et al. [23]. Agricultural, forested, wetland and bare ground categories were classified using MDP %IS coefficients [39].

The C-CAP data product is one source of land cover information incorporated into the NLCD dataset. However, C-CAP data only includes areas within the estuarine drainage area (Fig. 1) defined by National Oceanic and Atmospheric Administration (NOAA) National Ocean Service [9]. C-CAP data are scheduled for update at five year intervals and are currently available for 1991, 1996, 2001 and 2005 [9]. Data was processed using methods comparable to those described previously for NLCD [9, 14]. C-CAP pixels were assigned to one of 25 LULC categories (Table 2). LULC categories were reclassified to the mean %IS values for comparable NLCD 2001 categories [42] and MDP LULC categories (Table 2).

RESAC was contracted by the Chesapeake Bay Program to develop %IS estimates for the Chesapeake Bay watershed for the years 1990 and 2000 [42]. Several sources of imagery including Landsat 5 TM, Landsat 7 ETM+, IKONOS and orthophotos [43] were used. High resolution IKONOS imagery allowed for a sub-pixel regression tree classification process to estimate %IS from the Landsat imagery [43]. Percent IS validation was based on comparisons with IKONOS and orthophoto imagery from Montgomery County, Maryland and the surrounding Mid-Atlantic region. In the dataset available to the author, features having less than 10% IS had been assigned a value of zero.

Towson University was contracted by MDNR to develop %IS estimates for Maryland 12 digit watersheds (McGinty Margaret, MDNR, personal communication). Imagery used was from 1999 and 2000. Differences in development from 1999 to 2000 were assumed to be insignificant. Eight broad LULC categories were defined: deciduous, evergreen, herbaceous, impervious high, impervious low, bare ground and water. These broad categories are comparable to Anderson level I categorization. Tax data for the year 2000 was used for comparison with the TU %IS data.

NLCD, C-CAP and RESAC raster cells located within 12 digit watershed boundaries, excluding cells over open water, were extracted using the ArcGIS Extract by Mask tool. For each raster, the weighted mean %IS per 12 digit watershed [10] was calculated using the ArcGIS Zonal Statistics as Table tool. Twelve digit watershed %IS was already calculated for the TU data. All %IS data were joined to the corresponding year's tax data by 12 digit watershed.

2.2 Model Development

Global regression models describe the relationship between a dependent and independent variable(s) across the entire spatial extent of the data [44], which in this case is the State of Maryland. Ordinary least squares regression (OLS) was used to develop a global model to describe the relationship between total structure foundation square feet (ft^2) and %IS at the 12 digit watershed scale. Global regression models from OLS were expressed in the form

$$y_i = \beta_0 + \beta_{1x_i} + \varepsilon_i \qquad (1)$$

Where, y_i is the predicted %IS for 12 digit watershed i, β_1 is the slope, x_i is the observed tax data (ft^2/h) for 12 digit watershed i, β_0 is the y-intercept and ε_i is model error [44, 45].

When tax and %IS data are randomly distributed among 12 digit watersheds, the under-predictions and over-predictions (ε_i) from each model will be randomly distributed among the 12 digit watersheds. Non-random spatial distributions (spatial autocorrelation) violate two regression assumptions: (1) observations are independent of each other and (2) residuals are normally distributed with a mean of zero [46]. All OLS model residuals were tested for spatial autocorrelation with an inverse distance weighted Morans I statistic:

$$\mathrm{I} = \frac{n}{S_o} \frac{\sum_{i-1}^{n} \sum_{j-1}^{n} w_{ij} z_i z_j}{\sum_{i-1}^{n} z_i^2} \qquad (2)$$

Where, I is the Morans I statistic, n is the total number of 12 digit watersheds, w_{ij} is a weighting measure of spatial proximity between 12 digit watersheds i and j, z_i and z_j are the residual deviations of 12 digit watersheds i and j from the mean residual value and S_o is the aggregation of spatial weights [27]. The Morans I statistic determined if homogeneous clusters of over-predictions or under-predictions existed among the 12 digit watersheds.

If OLS regression models were significant and spatially autocorrelated then, geographically weighted regression (GWR) models were developed with the same dependent and independent variables. Geographically weighted regression is necessary to adjust the global OLS regression model for each 12 digit watershed [44-47]. A GWR is a variant of the linear OLS regression model, but without the assumption of randomly distributed error. It differs in that spatial bias in the data distribution is incorporated [47] into the regression model:

$$yi = \beta_0\left(u_i, v_i\right) + \sum_{j=1}^{k} x_{ij}\beta_j\left(u_i, v_i\right) + \varepsilon_i \qquad (3)$$

by only using a specified number of neighboring features. In GWR, y_i is the %IS (u_i,v_i) observation coordinates for 12 digit watershed (u,v), w_{ij} is the number of observations for each 12 digit watershed (i = 1,…,n) and its explanatory tax data (j = 1,…,k), β_0 and β_j are parameters for each 12 digit watershed and ε_i is model error for each 12 digit watershed i [44]. The result is a unique linear model for each 12 digit watershed. GWR uses a spatial kernel to determine which neighboring features to use when calculating the local regression equation [44, 45]. Spatial density of 12 digit watersheds varied, so, a variable kernel was used to determine which 12 digit watersheds were used for GWR models. Corrected Akaike Information Criterion (AICc) was used as the spatial kernel's bandwidth [45]. Model residuals were tested for spatial autocorrelation with the Morans I statistic previously described.

AIC_c and standard error were used to evaluate model performance among all OLS and GWR models [45]. The AIC_c kernel was used, instead of a cross validation kernel, because it allowed direct comparison of models having different sample sizes, different explanatory variables and different complexity. Fotheringham, Brunsdon, and Charlton [44] defined AIC_c as:

$$AIC_c = 2n\log_e\left(\hat{\sigma}\right) + n\log_e\left(2\pi\right) + n\left\{\frac{n+tr\left(S\right)}{n-2-tr\left(S\right)}\right\} \quad (4)$$

Where, n is the total number of 12 digit watersheds, $\hat{\sigma}$ is the residuals' standard error and tr(S) is the trace of the hat matrix for observed tax data and predicted %IS [46]. Models were considered different when the AIC_c differed by a value of four or more [46]. Preferred models are those having the lowest AIC_c value.

3. Results and Discussion

A total of 1,120 12 digit watersheds were used for the NLCD, RESAC and TU models, but only 901 of those watersheds were used for the C-CAP models (Fig. 1). The watersheds varied in size from 0.0066 hectares to 24,038 hectares. Watersheds that exceeded 6,070 ha typically had large expanses of open water.

Maryland tax data are spatially explicit at the land parcel level making it a direct measure of land development. The majority of tax records were spatially enabled with x, y coordinates, which allowed each record to be assigned to a single 12 digit watershed. Wu et al. [25] found that attributes related to a building's area (maximum area, standard deviation of building's area and the total building-area percentage) were among the most important factors for classifying land use. Not all Maryland land parcels had a primary structure (building), but for those that did, between 54% and 90% among the 24 counties were spatially enabled (Table 1). Not all of these records included the year the structure was built or its foundation area. None of the tax records included secondary structures such as: a detached garage, swimming pool, covered porch or paved surfaces [25]. The TU 2000 data had 23 records where %IS/h exceeded 40% when there was < 5,000 ft^2/h of building footprint (4.6% IS). The lack of metadata precludes an understanding of how these outliers were generated and so they were retained in the analysis.

Impervious surfaces are heterogeneously distributed in Maryland. The highest concentration of %IS occurs along the I-95 corridor between Baltimore City and Washington D.C. (Fig. 2a). Smaller urban centers are also visible. A large portion of Maryland has less than 5%IS. However, the dispersion and extent of impervious surfaces varies among the methods used to interpret the remotely sensed data (Fig. 2a).

Tax data was significantly correlated with %IS regardless of regression method (OLS global statewide versus GWR local 12 digit watershed) and resolution (Anderson level I classification versus Anderson level II classification). Increasing in ft^2/h were significantly related to increases in observed %IS (Table 3). OLS models accounted for significant amounts of index

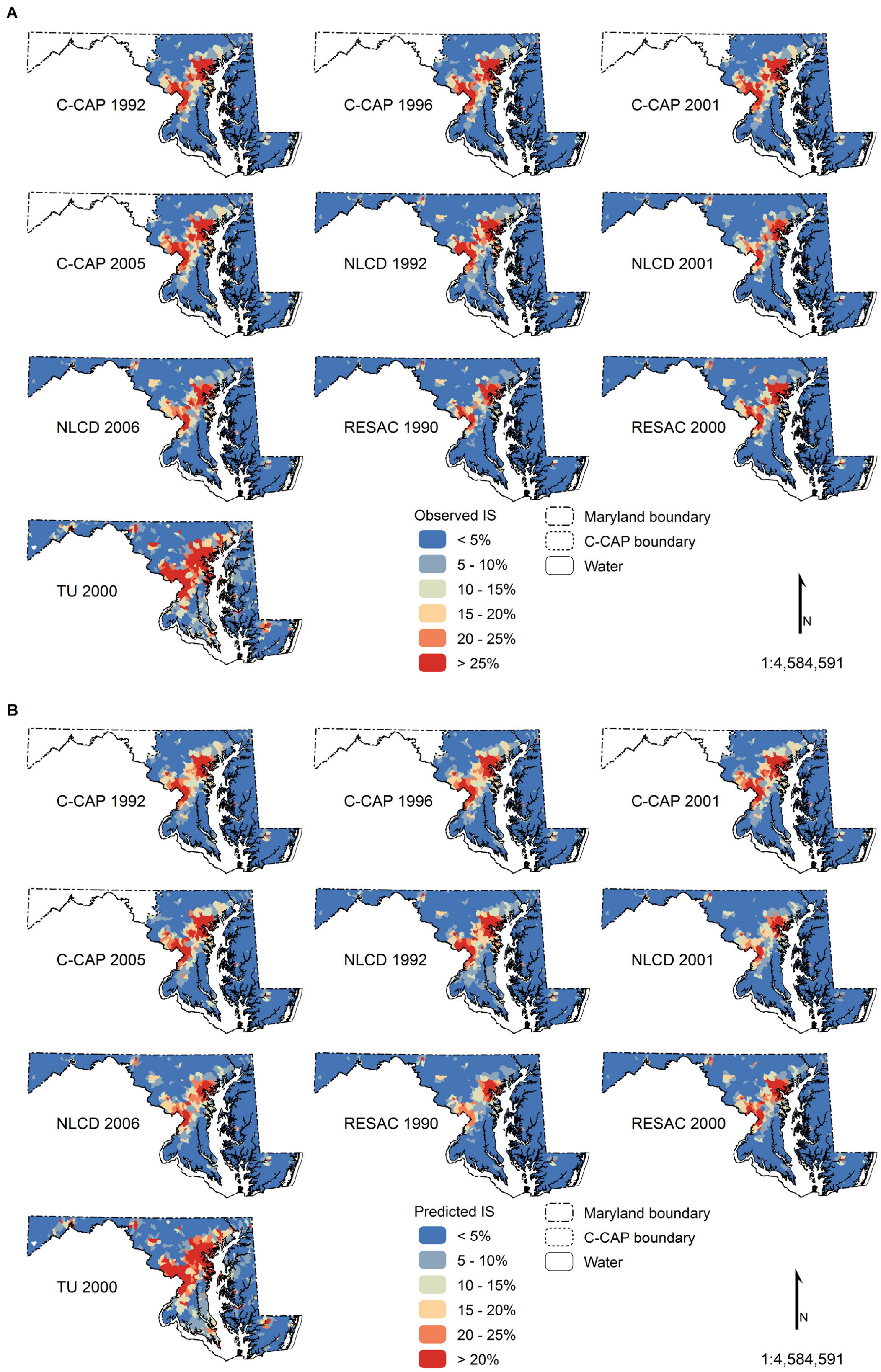

Fig. 2 Observed (a) and predicted (b) percent impervious surface distributions, GWR was used to generate the predictions.

Table 3 OLS equations for %IS versus foundation ft^2/h and Morans I spatial autocorrelation results for standardized residuals. Joint F-statistic indicates overall model significance and Jarque-Bera Statistic indicates if model residuals deviate from a normal distribution.

	Ordinary least squares regression							Morans I		
IS dataset	N	y	slope	R^2	AICc	Joint F-statistic	Jarque-bera statistic	Moran's index	Expected index	z-score
C-CAP 1992	864	3.05	0.00239	0.585	5,530	1,217*	44,951*	0.330	-0.00116	26.8*
C-CAP 1996	864	2.93	0.00239	0.595	5,526	1,271*	46,136*	0.322	-0.00116	26.2*
C-CAP 2001	864	1.56	0.00310	0.802	4,940	3,496*	30,686*	0.265	-0.00116	21.5*
C-CAP 2005	864	1.92	0.00245	0.813	4,920	3,742*	9,772*	0.330	-0.00116	26.5*
NLCD 1992	1,094	3.13	0.00232	0.530	6,949	1,233*	45,574*	0.336	-0.000915	30.4*
NLCD 2001	1,091	0.30	0.00255	0.760	5,894	3,455*	76,939*	0.253	-0.000917	23.0*
NLCD 2006	1,095	1.06	0.00204	0.809	5,802	4,634*	24,905*	0.281	-0.000914	25.4*
RESAC 1990	1,092	1.32	0.00213	0.560	6,566	1,385*	167,561*	0.280	-0.000917	25.7*
RESAC 2000	1,092	1.04	0.00286	0.755	6,131	3,360*	143,532*	0.264	-0.000917	24.2*
TU 2000	1,074	3.81	0.00529	0.683	7,749	2,307*	63,474*	0.260	-0.000932	23.4*

* Significant results ($p < 0.00001$).

variability: R^2 = 0.530 to 0.813 (Table 3). Model performance improved as %IS datasets became more recent. The C-CAP 2005 model performed best: R^2 = 0.813 and AIC_c = 4,920 (Table 3), but C-CAP does not provide coverage for all of Maryland. Using data for all of Maryland, the NLCD 2006 model was preferred: R^2 = 0.809 and AIC_c = 5,802 (Table 3). Predicted %IS slopes among the ten models had similar trajectories and %IS estimation errors. Each model suffered from %IS under prediction when the total structure footprint area was < 10,000 ft^2/h and over prediction > 30,000 ft^2/h of footprint (Fig. 3a). Jarque-Bera statistics, Morans I statistics (Table 3) and residual plots (Fig. 3b) confirmed that over-predictions and under predictions did not have normal spatial distributions even though the OLS models were significant. Spatial plots of the residuals show pronounced, homogenous clusters in urbanized areas and major transportation corridors (Fig. 4a). Watersheds having residuals > 2 are concentrated around Baltimore city, Washington D.C. and the I-95 corridor. Negative residuals < -2 are less common and typically north of Washington, D.C.

GWR produced 864 local models for C-CAP data, 1,095 for NLCD data, 1,092 for RESAC data and 1,074 for TU data. All GWR models significantly improved upon the OLS models' representation of the relationship between tax data (ft^2/h) and remotely sensed %IS data. Overall, GWR models accounted for a greater amount of error than did OLS models: R^2 = 0.917 to 0.965 and reduced the spatial autocorrelation of residuals for all models (Table 4). The AICc and Morans I results did not clearly indicate an optimal model. The C-CAP 2005 model had the highest R^2 (0.965), the lowest AIC_c (4,003) and a spatially random distribution of residuals (Table 4), but C-CAP data does not include all of Maryland. For both NLCD and RESAC models, more recent data accounted for greater amounts of model error, reduced AIC_c values, but a greater degree of spatially auto-correlated residuals (Table 4). The TU model had the highest AIC_c value, but it had a high R^2 of 0.934 and a spatially random residual distribution (Table 4). The GWR %IS predictions closely mirrored observed %IS

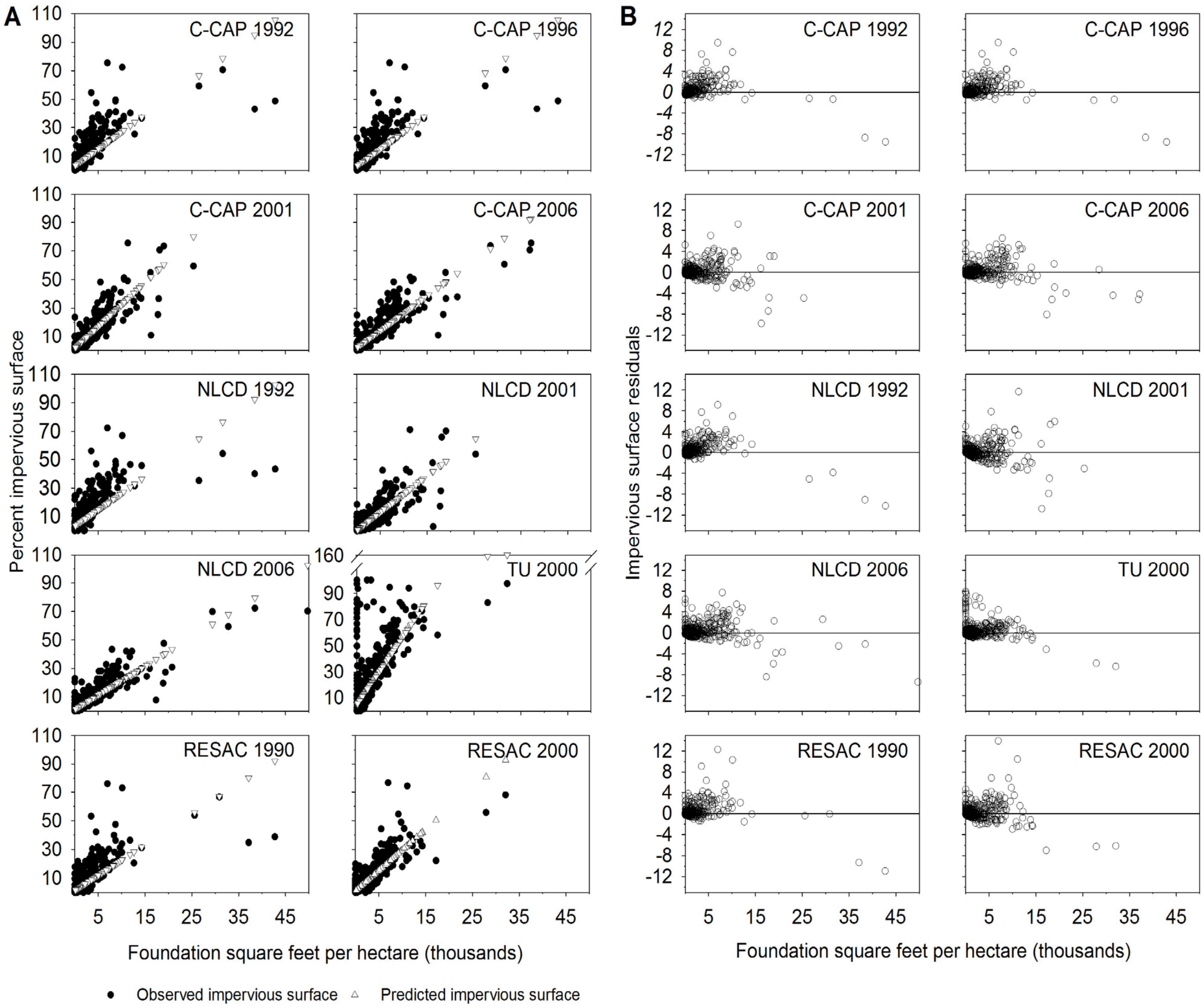

Fig. 3 (a) Predicted percent impervious surface as foundation area increased from each OLS model; (b) distribution of predicted impervious surface standardized residuals for the ten OLS models.

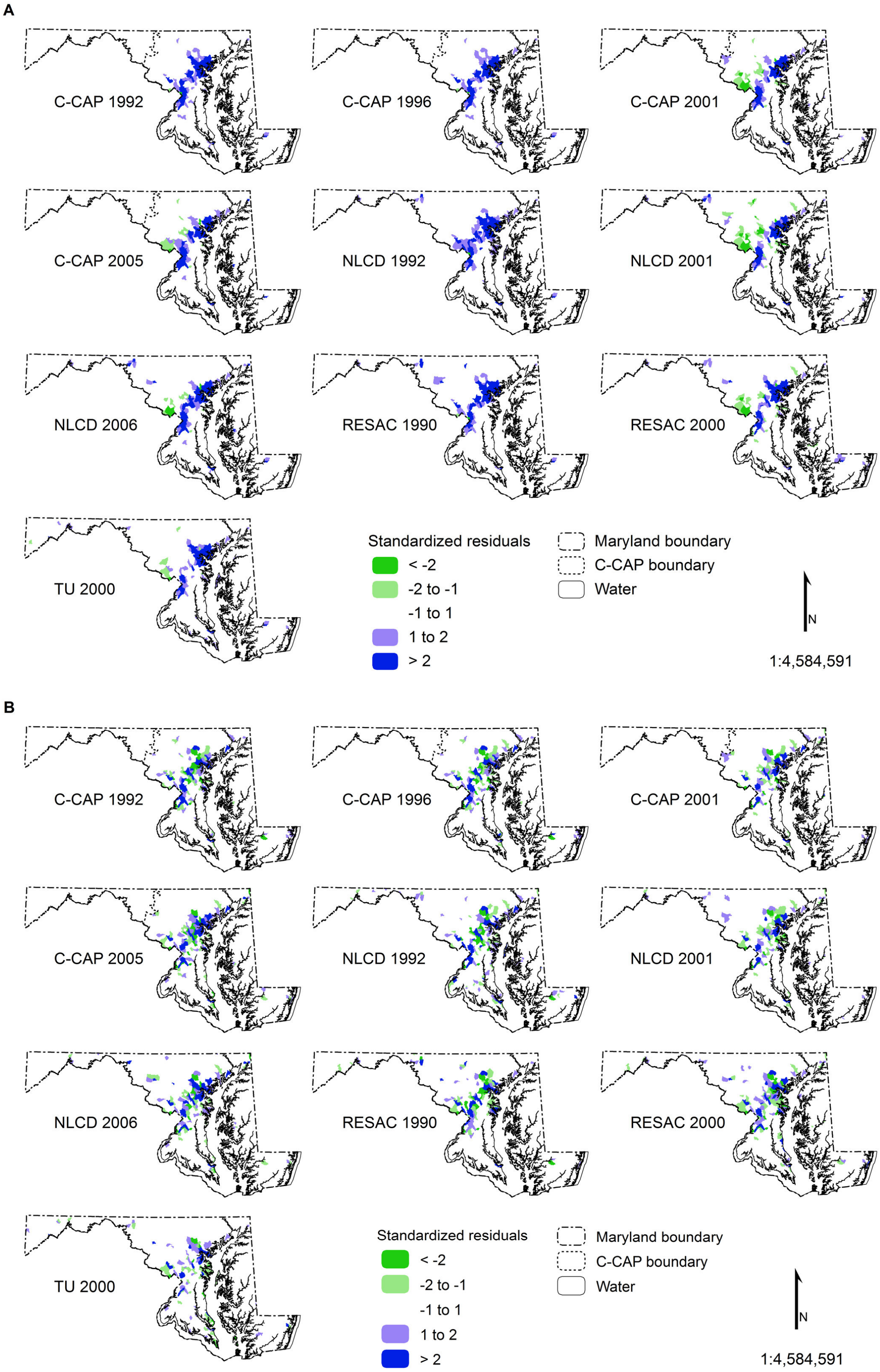

Fig. 4 **(a) Spatial distribution of model standardized residuals for OLS predictions and (b) GWR predictions.**

Table 4 GWR results for %IS versus foundation ft^2/h, Morans I spatial autocorrelation test results for the standardized residuals are included, asterisk indicates statistically significant results at $p \leq 0.05$.

IS dataset	Geographically weighted regression			Morans I			
	N	R^2	AICc	Moran's index	Expected index	z-score	p-value
C-CAP 1992	19	0.939	4,412	0.0281	-0.00116	2.37	0.018*
C-CAP 1996	19	0.939	4,428	0.0254	-0.000116	2.14	0.032*
C-CAP 2001	23	0.941	4,315	-0.0209	-0.00116	-1.61	0.11
C-CAP 2005	19	0.965	4,003	-0.00329	-0.00116	-0.171	0.86
NLCD 1992	19	0.934	5,458	0.0113	-0.000915	1.10	0.27
NLCD 2001	27	0.917	5,161	-0.0227	-0.000917	-2.00	0.045*
NLCD 2006	19	0.958	4,812	-0.0278	-0.000914	-2.42	0.015*
RESAC 1990	19	0.920	5,372	0.0139	-0.000917	1.35	0.18
RESAC 2000	13	0.963	5,194	-0.0273	-0.000917	-2.40	0.016*
TU 2000	16	0.934	6,881	-0.0176	-0.000932	-1.52	0.13

values (Fig. 5a) while, also reducing the Moran's I statistic (Table 4), removing spatial autocorrelation (Fig. 4b) and randomizing the distribution of residuals (Fig. 5b). GWR models performed poorly in portions of western Maryland, the upper eastern and lower western shores of Chesapeake Bay (Fig. 6a). Model performance for the Washington D.C. and Baltimore areas varied. Models not having spatially random residuals did have Morans I p-values that were considerably closer to 0.05 than did the OLS models (Tables 3 and 4) which was substantiated by the heterogeneous spatial distribution of positive and negative residuals (Fig. 4a). Percent IS standard errors among the GWR local models were relatively low except for TU: C-CAP 2005 was 0.132 to 2.01, NLDC 2006 was 0.112 to 1.81, RESAC 2000 was 0.0567 to 1.90 and TU 2000 was 0.228 to 4.67 (Fig. 6b).

As previously mentioned, MdProperty View tax data do not include paved surfaces such as roads, parking lots, driveways and other types of vehicle habitat which comprise vast amounts of IS. Chester, Horvath and Madanat [48] estimated parking density to range from 6.3-58 m^2 per 100 m^2 of road. This translates to 25-470 parking spaces per one kilometer of road. Furthermore, parking availability in metropolitan areas may be up to seven times the number of cars [5]. Goetz et al. [43] estimated that 36% of the IS in Maryland is due to roads, but state-wide %IS increased to 60% when all vehicle habitats were included.

While remote sensing data include all types of land surfaces, these data suffer from a lack of spatial resolution which limits their use to regional analyses and planning [41] and they are inappropriate for local-scale analyses and planning [14, 33]. Some %IS estimates are subject to considerable error, such as low-density and exurban development and are problematic [49]. The majority of algorithms developed to identify IS have been focused on urban lands. Many of those modeling techniques are unsuitable for rural lands [6, 8, 11, 25]. Bird et al. [23] compared direct measure of %IS to NLCD %IS estimates and demonstrated that the NLCD typically underestimated %IS for watersheds having 5% to 10%IS, which are aquatic ecosystem response thresholds [16]. Smith et al. [13] determined that NLCD level II thematic image classification accuracy was significantly affected by land cover heterogeneity and land cover patch size. Classification accuracy varied from 15%-25% for a ten pixel area to 52%-62% for a 10,000 pixel area. These are significant amounts of error considering that an appropriate resolution for local land planning is 4 m or less [33]. It is worth noting that McMahon [10] demonstrated having less %IS estimation error when using the more generalized NLCD level I

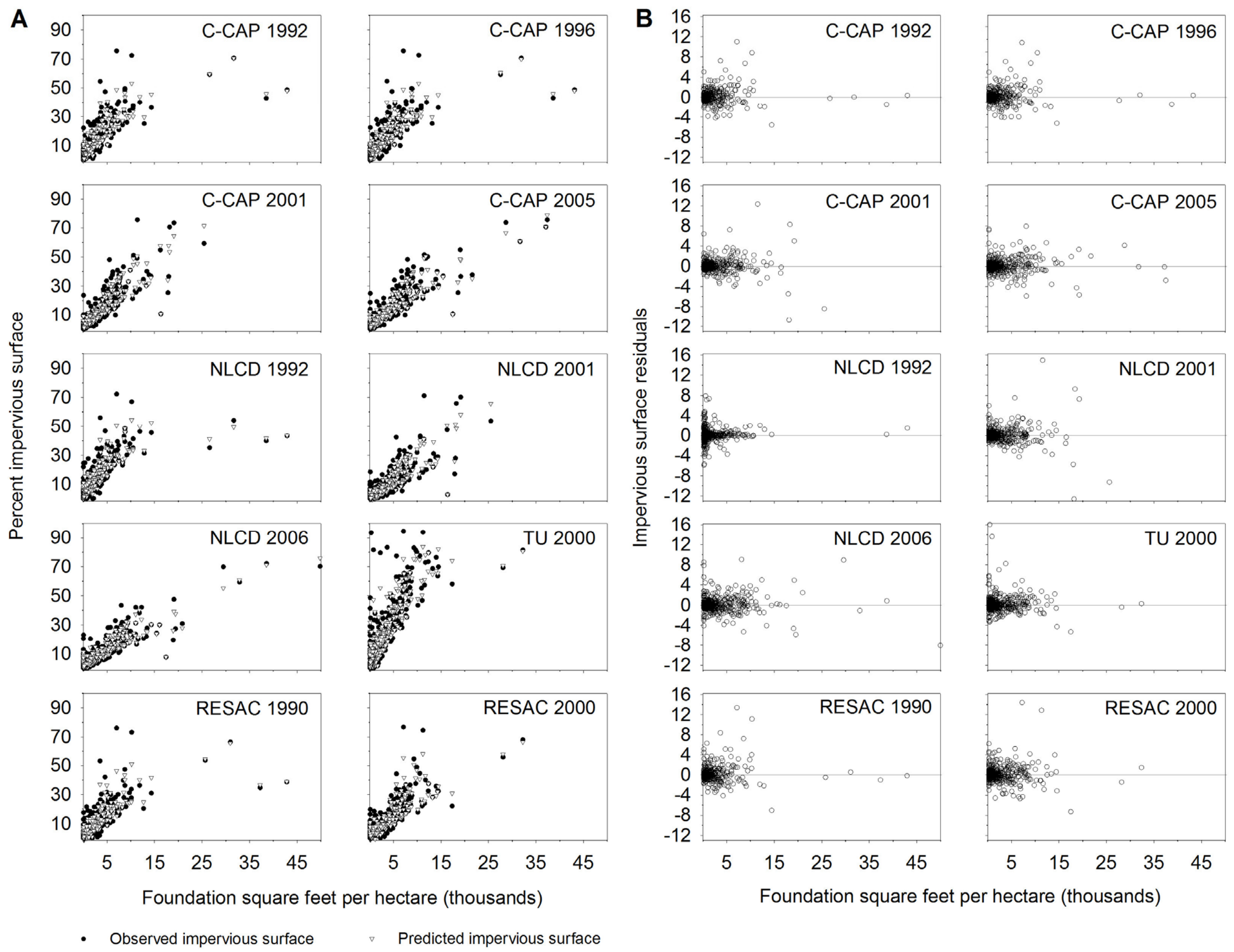

Fig. 5 (a) GWR predictions of percent impervious surface as foundation area increased and (b) the distribution of predicted impervious surface standardized residuals.

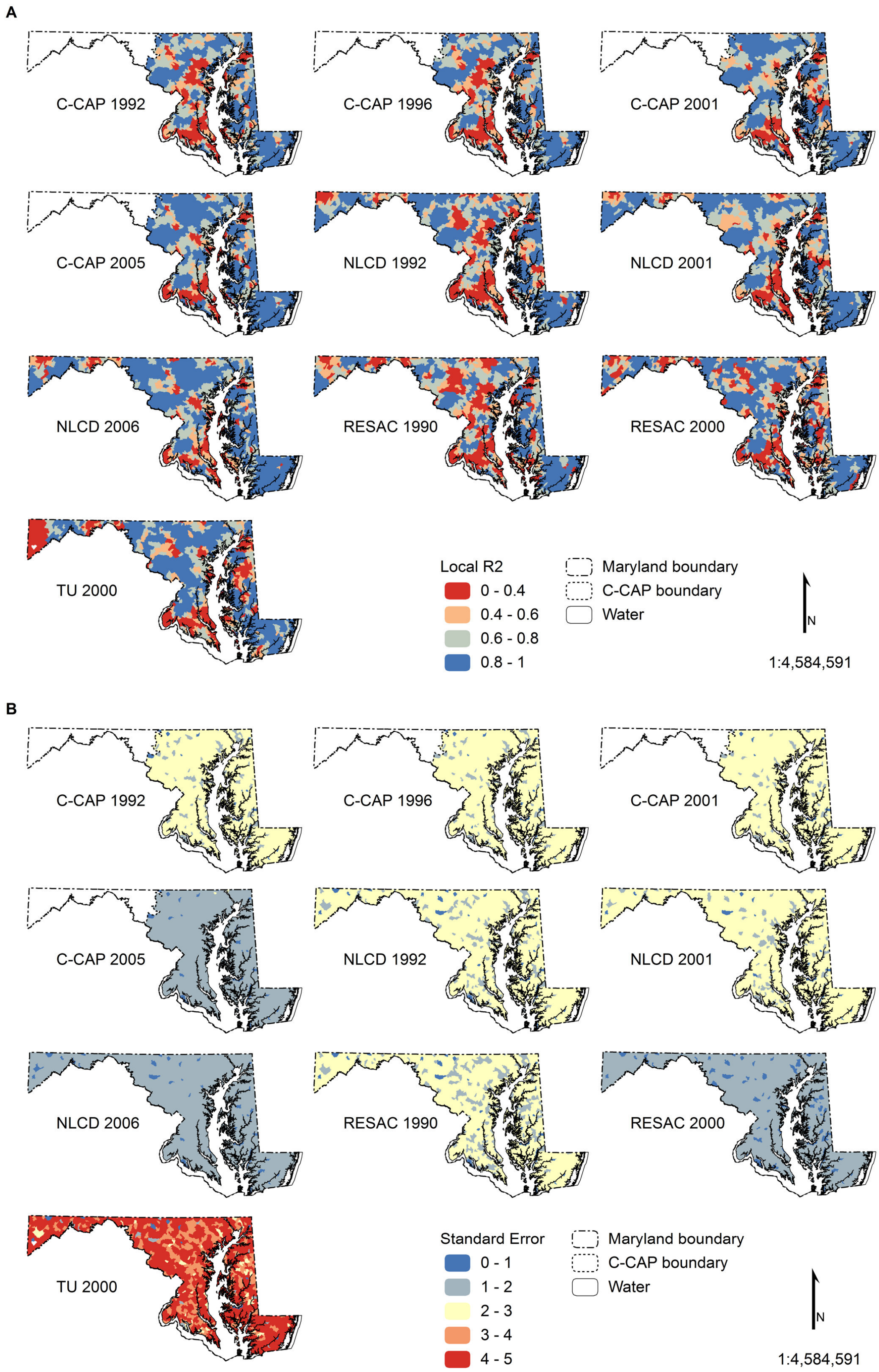

Fig. 6 (a) Spatial distribution of model performance (R^2) and (b) standard error for each GWR model.

land cover classification than when using the more detailed level II land cover classification. This suggests that the TU 2000 dataset having fewer categories and a steeper %IS response curve to ft^2/h may be a more realistic model.

There is a lack of effective IS modeling techniques and tools for rural land management. Consequently, the number of aquatic communities in watersheds at or exceeding ecological %IS reference points are likely underestimated. Annually indexing %IS estimates would facilitate use of established %IS reference points (5% and 10%) [16] to forecast ecosystem impairment under various land management scenarios. Watersheds having < 10% IS can be quickly identified as priority areas for conservation and protection [16]. For example, proposed development build-out scenarios (ft^2/h) can be quickly converted to a range of %IS estimates to predict the likelihood and extent of ecological impact. The simplicity of using tax data to estimate %IS would enable resources managers to evaluate potential ecological responses and engage with other stakeholders during the planning process, prior to ecosystem impact. Without a simple tool using current data, such as a tax map index, effective proactive and adaptive resource management will be difficult at best.

4. Conclusions

This analysis demonstrated that a strong positive relationship exists between the tax data metric (ft^2/h) and %IS estimates from remote sensing datasets. OLS regression produced significant global statewide tax index models. However, significant spatial clustering (spatial autocorrelation) was present among all models. Use of these models for local decision making within areas of spatial autocorrelation would not be advised. GWR is recommended to remove the spatial autocorrelation and associated error from the tax index. Used in concert, the various models presented here can provide a range of %IS outcomes based on various local land development scenarios.

Tax models index %IS as ft^2/h (or acre): a common measure of land development thereby facilitating effective communication among resource managers, land use planners, administrators, public officials and citizens. Effective communication highlights the need for a simple, easy to use index of land development. Unlike Reilly et al. [11], no data transformation was used in this analysis. Percent IS datasets are not updated annually causing a lag in detection of %IS trends to local planning and land development time frames. Tax data are standardized, annually updated, readily accessible and contain parcel and structure attributes which track both the addition and removal of IS on a parcel. It is important to emphasize that the tax based index does not generate an exact measure of %IS since paved surfaces are not included. A tax-based index allows rapid projection of future impacts for various development and build out scenarios. MdProperty View tax records date back to the 1700s. The opportunity exists to reconstruct time series of changing land conditions, development patterns and ecosystem responses.

An advantage of tax data is that parcel development records are independent of demographics. Use of demographic data to calculate %IS is not advised. Demographic metrics, such as job projections [11] and census data [23] have been proposed as %IS estimators. Demographic metrics are problematic because they fluctuate spatially unlike the spatial permanency of IS. Once structures and infrastructure are built they typically persist regardless of occupancy or use. A parallel scenario exists for areas having numerous vacation and seasonal residences or areas with commercial, industrial and mining activity [5]. Job availability is problematic because the watershed in which a job is located is not a predictor of the watershed in which a person will take up residence. Furthermore, demographic data does not capture parcel enhancement or redevelopment—the addition or removal of IS. Demographic metrics increase the risk of error propagation through the %IS calculation

[5]. Error propagation from census projections roughly doubles from the 10-year to 20-year forecasts. Exum et al. [5] recommend that census data was used for retrospective analyses rather than making projections.

Acknowledgements

Development and analysis of the tax data index benefitted from numerous comments and suggestions provided by Jim Uphoff and Margaret McGinty with the Maryland Department of Natural Resources. Jean Appleton with the Maryland Department of Planning provided much insight into the structure and content of the MdProperty View datasets. The author would like to thank the U. S. Fish and Wildlife Service for publishing support through federal aid to sport fishing.

References

[1] Jones, K. B., Neale, A. C., Nash, M. S., Van, Remortel, R. D., Wickham, J. D., and Riitters, K. H. et al. 2001. "Predicting Nutrient and Sediment Loadings to Streams from Landscape Metrics: A Multiple Watershed Study from the United States Mid-Atlantic Region." *Landscape Ecology* 16: 301-12.

[2] Lee, J. H., and Bang, K. W. 2000. "Characterization of Urban Stormwater Runoff." *Water Research* 34: 1773-80.

[3] 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: 243-58.

[4] Beach, D. 2002. *Coastal Sprawl: The Effects of Urban Design on Aquatic Ecosystems in the United States*. Arlington, VA: Pew Oceans Commission.

[5] Exum, L. R., Bird, S. L., Harrison, J., and Perkins, C. A. 2005. *Estimating and Projecting Impervious Cover in the Southeastern United States*. Athens, GA: Ecosystems Research Division, U.S. Environmental Protection Agency.

[6] Ji, M., and Jensen, J. R. 1999. "Effectiveness of Subpixel Analysis in Detecting and Quantifying Urban Imperviousness from Landsat Thematic Mapper Imagery." *Geocarto International* 14: 31-40.

[7] Cappiella, K., and Brown, K. 2001. *Impervious Cover and Land Use in the Chesapeake Bay Watershed*. Annapolis, MD: Chesapeake Bay Program, U.S. Environmental Protection Agency.

[8] Bauer, M. E., Heinert, N. J., Doyle, J. K., and Yuan, F. 2004. "Impervious Surface Mapping and Change Monitoring Using Landsat Remote Sensing." In *ASPRS Annual Conference Proceedings*, 10.

[9] Dobson, J. E., Bright, E. A., Ferguson, R. L., Field, D. W., Wood, L. L., and Haddad, K. D. et al. 1995. *NOAA Coastal Change Analysis Program* (*C-CAP*)*: Guidance for Regional Implementation*. U.S.: National Oceanic and Atmospheric Administration.

[10] McMahon, G. 2007. "Consequences of Land-Cover Misclassification in Models of Impervious Surface." *Photogrammetric Engineering and Remote Sensing* 73: 1343-53.

[11] Reilly, J., Maggio, P., and Karp, S. 2004. "A Model to Predict Impervious Surface for Regional and Municipal Land Use Planning Purposes." *Environmental Impact Assessment Review* 24: 363-82.

[12] Hu, X., and Weng, Q. 2011. "Estimating Impervious Surfaces from Medium Spatial Resolution Imagery: A Comparison between Fuzzy Classification and LSMA." *International Journal of Remote Sensing* 32: 5645-63.

[13] Smith, J. H., Wickham, J. D., Stehman, S. V., and Yang, L. 2002. "Impacts of Patch Size and Land-Cover Heterogeneity on Thematic Image Classification Accuracy." *Photogrammetric Engineering and Remote Sensing* 68: 65-70.

[14] Vogelman, J. E., Sohl, T. L., Campbell, P. V., and Shaw, D. M. 1998. "Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data Sources." *Environmental Monitoring and Assessment* 51: 415-28.

[15] Sokal, R. R., and Rohlf, F. J. 2000. *Biometry: The Principles and Practice of Statistics in Biological Research*. New York: W. H. Freeman and Company.

[16] Uphoff, J. H., McGinty, M., Lukacovic, R., Mowrer, J., and Pyle, B. 2001. "Impervious Surface, Summer Dissolved Oxygen, and Fish Distribution in Chesapeake Bay Subestuaries: Linking Watershed Development, Habitat Conditions, and Fisheries Management." *North American Journal of Fisheries Management* 31: 554-66. doi:10.1080/02755947.2011.598384.

[17] Stranko, S. A., Hilderbrand, R. H., Morgan, R. P., Staley, M. W., Becker, A. J., and Roseberry-Lincoln, A. et al. 2008. "Brook Trout Declines with Land Cover and Temperature Changes in Maryland." *North American Journal of Fisheries Management* 28: 1223-32. doi:10.1577/M07-032.1.

[18] Price, S. J., Cecala, K. K., Browne, R. A., and Dorcas, M. E. 2010. "Effects of Urbanization on Occupancy of Stream Salamanders." *Conservation Biology* 25: 547-55. doi:10.1111/j.1523-1739.2010.01627.x.

[19] Center for Watershed Protection. 2003. *Impacts of Impervious Cover on Aquatic Systems*. Ellicott City: MD, Center for Watershed Protection.

[20] Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Arlington, VA: Geological Survey, United States Department of the Interior.

[21] Stehman, S. V., Wickham, J. D., Smith, J. H., and Yang, L. 2003. "Thematic Accuracy of the 1992 National Land-Cover Data for the Eastern United States: Statistical Methodology and Regional Results." *Remote Sensing of Environment* 86: 500-16.

[22] Smith, J. H., Stehman, S. V., Wickham, J. D., and Yang, L. 2003. "Effects of Landscape Characteristics on Land-Cover Class Accuracy." *Remote Sensing of Environment* 84: 342-9.

[23] Bird, S., Harrison, J., Exum, L., Alberty, S., and Perkins, C. 2002. "Screening to Identify and Prevent Urban Storm Water Problems: Estimating Impervious Area Accurately and Inexpensively." In *Proceedings of the National Water Quality Monitoring Council Conference*, 9-13.

[24] Moglen, G. E., and Beighley, R. E. 2002. "Spatially Explicit Hydrologic Modeling of Land Use Change." *Journal of the American Water Resources Association* 38: 241-52.

[25] Wu, S., Silvánhyphen, J., and Wang, L. 2007. "Per-Field Urban Land Use Classification Based on Tax Parcel Boundaries." *International Journal of Remote Sensing* 28: 2777-800.

[26] Stone, B. 2004. "Paving Over Paradise: How Land Use Regulations Promote Residential Imperviousness." *Landscape and Urban Planning* 69: 101-13.

[27] Environmental Systems Resource Institute [ESRI]. 2012. *ArcGIS Desktop* (*Version 10.1*). Redlands, CA: ESRI.

[28] Maryland Department of Planning. 2009. *MdProperty View Subscriber's Guide*. Baltimore: MD, Maryland Department of Planning.

[29] Maryland Department of Natural Resources. 1998. "SWSHED12." MD IMAP: Maryland's Mapping and GIS Data Portal. Accessed December 1, 2008. http://data.imap.maryland.gov/datasets/1f19823dd037425dabfd0f45724814be_2.

[30] Strahler, A. N. 1957. "Quantitative Analysis of Watershed Geomorphology." *Transactions, American Geophysical Union* 38: 913-20.

[31] Fry, J. A., Coan, M. J., Homer, C. G., Meyer, D. K., and Wickham, J. D. 2009. *Completion of the National Land Cover Database (NLCD) 1992-2001 Land Cover Change Retrofit Product*. Reston, VA: Geological Survey.

[32] Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., and Wickham, J. 2007. "Completion of the 2001 National Land Cover Database for the Conterminous United States." *Photogrammetric Engineering and Remote Sensing* 73: 337-41.

[33] Vogelman, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., and Van, Driel, N. 2001. "Completion of the 1990s National Land Cover Data Set for the Conterminous United States for Landsat Thematic Mapper Data and Ancillary Data Sources." *Photogrammetric Engineering and Remote Sensing* 67: 650-5.

[34] NOAA Coastal Services Center. 2009. "Coastal Change Analysis Program Regional Land Cover and Change: Coastal Change Analysis Program (C-CAP) Classification Scheme and Class Definitions." NOAA Office for Coastal Management, National Oceanic and Atmospheric Administration. Accessed October 19, 2011. http://coast.noaa.gov/digitalcoast/_/pdf/ccap_class_scheme.pdf.

[35] University of Maryland, College Park, Maryland State Highway Administration, Mid-Atlantic Regional Earth Science Applications Center (RESAC). 2003. *RESAC 2000 CBW Impervious Surface Product, Version 1.3*. College Park, MD: Chesapeake Bay Program, U.S. Environmental Protection Agency.

[36] University of Maryland, College Park, Mid-Atlantic Regional Earth Science Applications Center (RESAC). 2003. *RESAC 1990 CBW Impervious Surface Product, Version 1.3*. College Park, MD: Chesapeake Bay Program, U.S. Environmental Protection Agency.

[37] Barnes, K. B., Morgan, J. M., Roberge, M. C., and Lowe, S. *Sprawl Development: Its Patterns, Consequences, and Measurement*. Towson: Center for Geographic Information Sciences.

[38] Geological Survey EROS Data Center (EDC). 2000. "National Land Cover." Sioux Falls, SD: Geological Survey. Accessed September 22, 2011. http://www.mrlc.gov/finddata.php.

[39] Center for Watershed Protection. 2005. *A User's Guide to Watershed Planning in Maryland*. Annapolis, MD: Watershed Services, Maryland Department of Natural Resources.

[40] Geological Survey EROS Data Center (EDC). 2011. "National Land Cover Database." Multi-Resolution Land Characteristics Consortium (MRLC). Geological Survey. Accessed September 27, 2011. http://www.mrlc.gov/.

[41] Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M. 2004. "Development of a 2001 National Landcover Database for the United States." *Photogrammetric Engineering and Remote Sensing* 70: 829-40.

[42] Geological Survey EROS Data Center (EDC). 2011. "National Land Cover Dataset 1992 (NLCD1992): Product Legend." Accessed October 19, 2011. http://www.mrlc.gov/nlcd92_leg.php.

[43] Goetz, S. J., Jantz, C. A., Prince, S. D., Smith, A. J., Varlyguin, D., and Wright, R. K. 2004. "Integrated

Analysis of Ecosystem Interactions with Land Use Change: The Chesapeake Bay Watershed." In *Ecosystems and Land Use Change*, edited by DeFries, R. S., Asner, G. P., and Houghton, R. A. Washington, DC: American Geophysical Union, 263-75.

[44] Fotheringham, A. S., Brunsdon, C., and Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England: John Wiley & Sons Inc..

[45] Fotheringham, A. S., Brunsdon, C., and Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England: John Wiley & Sons Inc..

[46] Charlton, M., and Fotheringham, A. S. 2009. "Geographically Weighted Regression White Paper." Geographically Weighted Regression. Accessed October 24, 2011. http://gwr.nuim.ie/downloads/GWR_WhitePaper.pdf.

[47] Mitchell, A. 2005. *The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics*. Redlands, CA: Environmental Systems Resource Institute (ESRI).

[48] Chester, M., Horvath, A., and Madanat, S. 2010. "Parking Infrastructure: Energy, Emissions and Automobile Life-Cycle Environmental Accounting." *Environmental Research Letters* 5: 1-8. doi:10.1088/1748-9326/5/3/034001.

[49] Irwin, E. G., and Bockstael, N. E. 2007. "The Evolution of Urban Sprawl: Evidence of Spatial Heterogeneity and Increasing Land Fragmentation." In *Proceedings of the National Academy of Sciences* 104 (20): 672-7.