THE MID-ATLANTIC REGIONAL EARTH SCIENCE APPLICATIONS CENTER (RESAC): AN OVERVIEW

Scott J. Goetz, Stephen D. Prince, Michelle M. Thawley,
Andrew J. Smith, Robb Wright
Department of Geography
University of Maryland
College Park, MD 20742-8225

ABSTRACT

The mid-Atlantic Regional Earth Science Applications Center (RESAC) was established in the Geography Department at the University of Maryland (UMD) by NASA's Earth Science Applications Program. The mid-Atlantic RESAC is to provide improved land cover mapping and ecological modeling capabilities for a diverse consortium of partners in Government, Academia, Industry and NGOs within the 178,000 km^2 Chesapeake Bay watershed. It is one of 7 regional centers established nationwide and leverages expertise in satellite remote sensing to address applications of regional significance including land cover change, land use planning, carbon exchange modeling, and integrated environmental monitoring. Examples of issues that are being addressed include nutrient runoff to the Chesapeake Bay, urban sprawl, farm and forest productivity, landscape fragmentation effects on biodiversity, a land manager decision support system, and educational outreach. The mid-Atlantic RESAC provides an example of how scientific advances can be focused on practical applications that challenge our ability to manage resources sustainably.

A brief overview of the RESAC is provided and specific applications are reviewed using examples that emphasize the utility of remote sensing and GIS capabilities. Results of field activities undertaken during the 1999 growing season, for example, are used with a fusion of multi-temporal Landsat-7 Enhanced Thematic Mapper and SPOT panchromatic imagery to classify vegetation types, and to characterize development of the severe drought that took place in the region.

THE MID-ATLANTIC RESAC

The University of Maryland Department of Geography at College Park houses the mid-Atlantic Regional Earth Science Applications Center (RESAC), one of seven regional centers established to apply NASA products to a wide range of earth science and resource management issues. Foremost among these are land cover properties derived from digital remote sensing and spatial information technologies made possible by NASA research, including data from a variety of revolutionary new aircraft and satellite sensors. To accomplish this in the mid-Atlantic region, which encompasses the entire Chesapeake Bay watershed, the RESAC has partnered with nearly 40 educational, research, non-profit and governmental organizations at local, state and federal levels whose needs for this information are guiding application development. The ultimate goal of the RESAC is and to aid decision making on a wide range of issues including land use planning, agricultural practices, and forest and water management, and to provide improved information and technologies to assess regional impacts of environmental change.

We provide a brief overview of the mid-Atlantic RESAC, focused on land cover change, ecosystem modeling, land use planning, and technology applications. We then focus on a key objective - mapping land cover using multi-temporal Landsat-7 imagery, SPOT panchromatic imagery, and field sampling, all conducted in a geographical information system framework.

Land Cover Change

Among the main components of land use decisions is the condition of the land itself and changes in land use, in part revealed by land cover data obtainable through satellite remote sensing. In consultation with project partners, which include the Chesapeake Bay Program, the Mid-Atlantic RESAC has identified problem areas that existing land cover data is inadequate to address and that have high priority for decision-making purposes. For example, existing land cover data, derived from 30 meter resolution Landsat images, has not been adequate for monitoring of riparian buffer strips, which are important in controlling chemical runoff to the Chesapeake Bay. Other key challenges are to distinguish and map crop lands from grass and pasture lands, different densities of residential development, monitoring of forest growth and carbon dioxide exchange with the atmosphere, wetland loss, and conversion of farmland and forest areas to residential and commercial development.

Ecosystem Modeling

One purpose of developing these applications is to be able to provide better estimates of forest and crop production, as well as nutrient inputs to the Chesapeake Bay, including their sources and pathways. This information is needed for ecosystem modeling that is used to monitor the health of the Bay, and to enforce regulations for long-term restoration. This is a policy priority for the Chesapeake Bay Program, and is expected to become even more important this year, as caps on nutrient levels go into effect, because it will require new sources to be offset by reductions in other areas. Present modeling efforts tell us about overall trends but have been not been able to link specific management actions to their consequences for water quality and the living resources of the Bay.

Land Use Planning

Another application, not unrelated to restoration of the Bay, is to provide better information for planning and management of growth, consistent with the capacities of local and state governments as well as the environment. This is particularly important to provide the services needed to support a growing population. Monitoring urban growth in the Washington DC metropolitan area, based on changes observed in satellite imagery, shows a loss of 22 km^2 a year between 1973 and 1996, or approximately 600 square meters of land conversion per person. More detailed analysis of this kind of data, at higher resolutions, is expected to provide the basis for modeling that identifies specific areas where land conversion is expected, and provides the basis for analyzing trade-offs and establishing land use priorities.

Technology Applications

Initial RESAC activities have been directed at land cover mapping, in some cases to find out more precisely what can be accomplished with emerging new technologies. Among the advantages of Landsat 7, launched in April 1999 with the Enhanced Thematic Mapper (ETM+), is that the data are being made available very soon after acquisition, and at a more affordable cost to users. This has made it more feasible to use multi-temporal imagery which provides more information than can be obtained from one image at a single point in time and has important implications for land cover mapping and land use classification. This capability is particularly useful for distinguishing different types of crops and vegetation. It also makes it possible to work with more recent images when conducting fieldwork. In the summer of 1999 RESAC members conducted fieldwork throughout the Mid-Atlantic region to use in supervising a land cover classification and to provide independent validation of the utility of satellite data for mapping these important land cover distinctions. Although most Landsat data is 30 meters in resolution, a feature new to Landsat 7 is a 15 meter channel, which permits detection of finer scale features such as field boundaries and riparian buffers. The new IKONOS satellite can provide one meter data, but is expensive to use over large areas in a monitoring framework. Use of higher resolution images in test areas, however, provides another data source with which to validate features observed in lower resolution imagery. We demonstrate the utility of some of these high-resolution data sources using aircraft imagery recently acquired in the region.

IMPROVED LAND COVER MAPPING: A PRIMARY RESAC OBJECTIVE

A common thread in many of the RESAC partner activities is the need for improved land cover characterization. Given the breadth of activities within the Chesapeake Bay research community it is unreasonable to expect any single classification to adequately address many diverse land cover needs. In lieu of producing a generalized land cover map to be incorporated into various applications, the RESAC is developing methodologies where customized land cover products can be derived from a data base of remotely sensed and GIS data. This requires algorithms that are flexible in design and easily automated to remove the overhead involved in a manual process. The goal is to provide a database of land cover information and a mapping algorithm that is repeatable and modifiable to tailor to specific data requirements, ideally improving accuracy for the end user. A rich archive of supporting data are being accumulated in a GIS, including high spatial and temporal resolution remotely sensed data, hyperspectral data, census data, elevation data, digital line graphs, digital ortho photo quarter quadrangles, aerial photography, detailed county level planametric maps, field measurements, and NASS crop statistics. This archive serves as the basis for all land use land cover mapping exercises.

We tested with an initial goal of producing a Landsat 7 ETM+ land cover classification of the Chesapeake Bay watershed. Approximately 18 Landsat WRS-2 scenes were required to produce a complete image of the 178,000 km^2 watershed. The methodology was first applied to a single Landsat path/row, WRS-2 path 15, row 33, centered around the Washington, DC area. The resulting land cover map of the DC area, an accuracy assessment, and a multi-temporal analysis were completed.

Methods

A decision tree approach was chosen because of its ability to incorporate both remotely sensed data as well as categorical data. Trees can be pruned, grafted, or forced to split based on user defined criterion. In the hierarchical tree structure, each split in the tree results in two branches. The algorithm searches for the dependent variable, that if split into two groups, would explain the largest proportion of deviation of the independent variable. At each new split in the tree, the same exercise is conducted and the tree is grown until it reaches terminal nodes, or leaves, each node representing a unique training pixel. Every leaf has a land cover class assignment. The machine learning software package c5.0 was used to generate the classification trees (Quinlan 1993).

The dependent variables used to grow the tree include a multi-temporal time sequence of Landsat 5 data. Six dates were available: Mar 27, Apr 12, Apr 28, Jul 02, Aug 08 and Nov 22, for 1998. As more recent Landsat 7 data become available they will be included in the analysis. All bands except the thermal band 6 were included in the analysis. Other dependent data layers included a 30 meter DEM and corresponding slope and aspect, 1990 population data, and a physiographic provinces map.

Training sites were selected for each Anderson Level II land use class (Anderson 1976). Training data were derived from Landsat imagery enhanced by spectral merging with higher resolution panchromatic data from SPOT and ETM+. Spectral merging improves visual interpretation by compressing more spectral information into a visible range, which in turn improved the ability to identify sites for training. Agricultural and forest training data were collected during a month-long field sampling campaign conducted during the summer of 1999 using laptop computers and GPSs. Ancillary GIS vector coverages, digital ortho photo quarter quads, National Wetlands Inventory data, and other classification products were used to assist in site identification.

Previous research suggests that multi-temporal imagery is useful for discriminating spectrally similar land cover classes such as parklands, lawns, crops and pastures (Vogelmann et al. 1998). Given the increase in data availability with two Landsat satellites in orbit, it is useful to investigate the merits of multi-temporal versus a reduced frequency data set. The same training data were used to produce a classification of a single date (August 8^{th}) and also a leaf-on-leaf-off classification (March 27^{th} and August 8^{th}).

In order to evaluate the overall accuracy of the results as well as analyze the merits of multi-temporal data, two comparisons were conducted. The first comparison was with the USGS's Landsat based MRLC product for Federal region 3 (Vogelmann et al. 1998). The second comparison was with a vector based land cover map of the state of Maryland produced by the Maryland Office of Planning (www.op.state.md.us), derived manually from a variety of GIS based data and aerial photography. The MOP data cover only for the state of Maryland, approximately one-half of the DC scene.

Results

A land cover classification of the region centered on Washington, D.C. is shown in Figure 1. The resulting images were run through a 3-by-3 pixel sieve filter to reduce speckle. The first few splits in the classification tree are shown in Figure 2. The first split in the tree was based on April 12^{th} band 5. The next two binary splits occur with March 27^{th} band 2 and November 11^{th} band 1. This is interesting from a multi-temporal point of view because all of these dates would likely be omitted from a leaf-on / leaf-off analysis, as was done with MRLC. Spring and fall dates have been reported to be useful for increasing the accuracy of crops, pastures and grasslands (Vogelmann 1998).

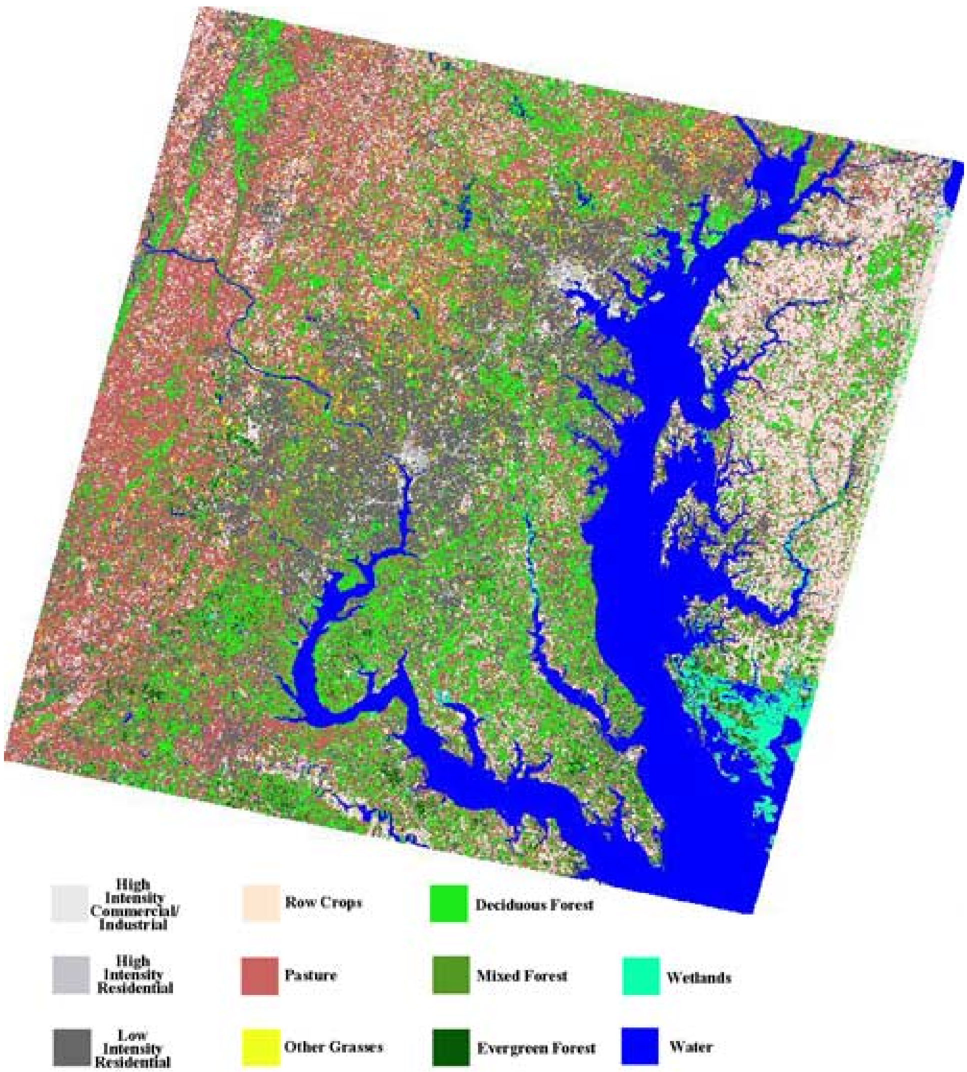

Figure 1. Land cover classification of Landsat path 15 row 33.

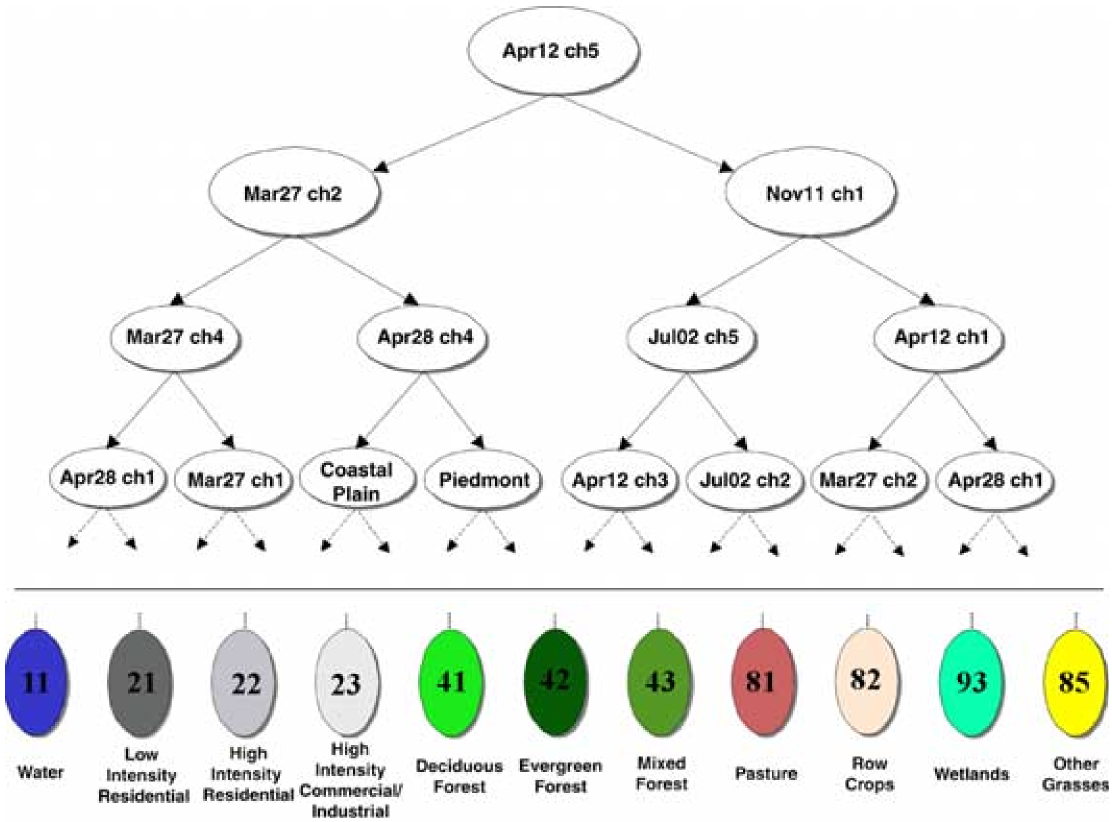

Figure 2. First few splits in decision tree classification.

Error matrices for the MRLC/RESAC and MOP/RESAC comparisons, as well as a comparison of MRLC/MOP for reference, are shown in Table 1a-c. The Kappa statistic (Hudson 1987), a measure of agreement between two matrices, was calculated for each of the three temporal analyses in terms of the two reference data sets. Kappa values are included in Table 1. Kappa values are all large and negative suggesting negative correlation between the reference data, both MRLC and MOP data, and current results. The comparison between MRLC and MOP data also portrays a poor relationship. Whereas a near zero kappa value suggests no relationship between two sources of data, a large negative value may be indicative of a systematic mislabeling of pixels. It is possible that interpretation of class definitions as described by Anderson (1976) could cause misclassification. This theory is supported by the fact that misclassification occurs primary between related land cover classes. For example many pixels labeled high density residential are labeled high density industrial/commercial in the reference data set.

(a) RESAC / MOP Comparison (Kappa=-0.56)

MOP \ RESAC	21	22	23	41	42	43	81	82	85	93
21	663052	580	16678	232089	13798	63819	620034	90861	100462	6805
22	1105970	22881	45286	27259	1956	13721	100431	16179	37566	1416
23	401941	28658	245495	20472	1880	6236	116725	50786	20938	2805
41	767125	2892	40018	1959651	57648	419239	1002844	239191	38323	33346
42	51053	213	4703	53861	115073	51741	34838	10585	1164	52163
43	281294	1023	15445	510266	128756	299087	212263	72318	6757	31621
81	83383	281	12020	34737	3747	8608	389486	104548	24137	2008
82	617174	2063	120443	152855	18470	47631	2196601	3088593	128388	20965
85	64045	1372	10817	10346	446	2943	56151	12540	52778	674
93	47214	2695	19485	28945	6044	13030	21126	14218	1594	312479

(b) RESAC / MRLC Comparison (Kappa=-0.90)

MRLC \ RESAC	21	22	23	41	42	43	81	82	85	93
21	888327	32270	118885	40369	8350	20806	205654	98924	90175	2228
22	66592	30360	38801	250	43	109	2467	2839	1443	390
23	67829	26112	88259	1128	323	678	10038	15856	1755	4052
41	942442	7685	98273	1520970	76691	438471	1171194	462280	114279	11400
42	133627	2022	19086	128265	170395	194452	98996	32997	22273	12950
43	272058	2223	26271	353897	63831	230212	258207	72575	48327	6591
81	816323	10739	196420	204241	19733	69055	2181349	1978596	327023	10286
82	393772	15004	155516	51539	5353	11812	500132	1039028	47188	7788
85	29187	1777	16052	1457	164	493	28036	11766	31915	125
93	151190	7293	46523	308514	90727	182991	151030	74477	24518	346496

(c) MOP / MRLC Comparison (Kappa=-0.74)

MRLC \ MOP	21	22	23	41	42	43	81	82	85	93
21	69979	59448	186	4193	196	1128	316	956	3565	56
22	15925	140854	687	10776	973	4376	1866	22067	5502	2505
23	0	0	0	0	0	0	0	0	0	0
41	187743	79727	6581	2487839	52629	564237	106886	677963	34368	18191
42	28396	9498	930	187308	128645	272457	12254	67867	4665	9454
43	70224	18385	1232	518387	41398	302735	29775	152243	9430	8320
81	142050	120660	6817	553740	18290	112880	406682	3674985	48634	16385
82	102206	120780	7958	149427	9283	40492	64676	1527262	14761	11421
85	17244	40316	685	10073	132	868	1237	4290	41568	695
93	8501	14088	2739	497678	110931	200466	14284	113492	2960	364019

Table 1a-c. Error matrices for comparison bewtween RESAC/MOP, RESAC/MRLC, and MOP/MRLC.

An error matrix for each of the three multi-temporal results, as compared to MRLC, is shown in Figure 3. Results suggest that high frequency data do improve the accuracy of most classes, primarily vegetated classes. For other classes the increase in accuracy is negligible. Urban classes for example do not benefit much from the multi-temporal data. Because of the flexibility of the decision tree design, this information can be incorporated into the tree design of future runs to improve overall accuracy. Until a more thorough validation can be conducted, however, these results are inconclusive.

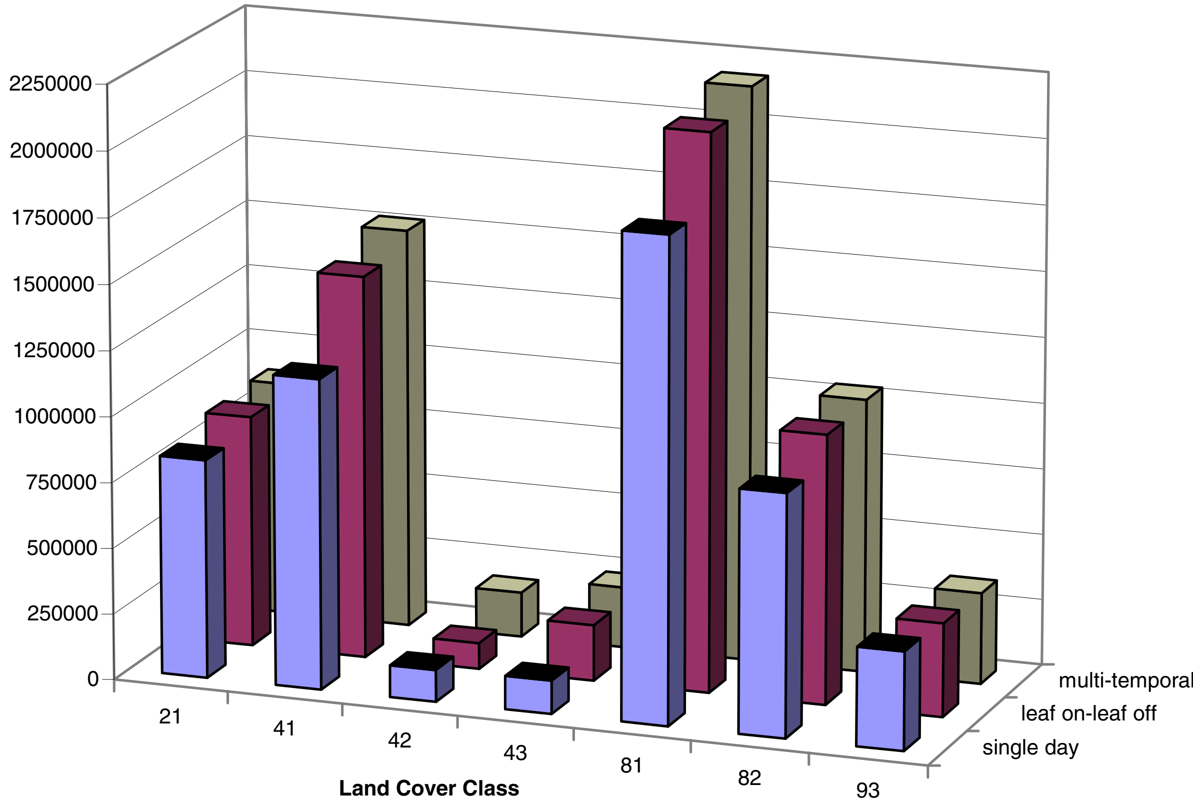

Figure 3. Bars represent pixel counts for each land cover class for each of the three temporal analyses.

CONCLUSIONS

There is significant disagreement between comparisons of three classification maps of the mid-Atlantic region surrounding Washington, D.C. One difficulty encountered during this research was identifying sources of data that could be considered "truth" and used in a validation exercise. This problem will be an even greater issue at the watershed scale. More research into validation methods, including in site field validation, or the use of county level GIS inventory and planning data, need to be investigated. Our next step will be to apply this methodology to each scene in the mid-Atlantic region, encompassing the Chesapeake Bay watershed, and merge the results to create a basin-wide mosaic. This classification is in progress as ETM+ data become more generally available.

While useful for understanding environmental patterns and conditions over a broad area such the Chesapeake Bay watershed, land cover information by itself is not sufficient for making land use decisions. Such decisions are inherently controversial because of multiple and often conflicting social objectives. They require information that links land conditions to causal factors and that helps to identify uncertainties, decision options, their implications for different stakeholders, and also anticipates conflicts. Thus, the Mid-Atlantic RESAC is developing an integrated monitoring approach and a Land Management Information System that can help prioritize information needs and better inform processes of negotiation and conflict resolution. This will require more extensive outreach activities, as this knowledge is in large part derived from the perceptions and experiences of land use planners, agricultural extension agents, citizen groups, land managers, and other stakeholders. The mid-Atlantic RESAC expects to expand partnerships with regional stakeholders as information is distributed through the NASA/UMD Earth Science Information Partnership. For more information see www.inform.umd.edu/geog/landcover/resac/.

REFERENCES

Anderson, J.R., E.E. Hardy, J.T. Roach, R.E. Witmer (1976) A land use and land cover classification system for use with remote sensor data, U.S. Geological Survey Professional Paper 964, 28pp.

Borak, J.S., A.H. Strahler (1999) Feature selection and land cover classification of a MODIS-like data set for a semiarid environment, In: *International Journal of Remote Sensing*, Vol.20(5), 919-938.

Vogelmann, J.E., T. Sohl, S.M. Howard (1998) Regional Characterization of land cover using multiple sources of data, In: *Photogrammetric Engineering and Remote Sensing*, Vol.64(1), 45-57.

Hudson, L.D., C.W. Ramm (1987) Correct formulation of the kappa coefficient of agreement, In: *Photogrammetric Engineering and Remote Sensing*, Vol.53(4), 421-422.

Quinlan, R. (1993) C4.5: Programs for Machine Learning. San Meteo, CA: Morgan Kaufmann.