

A Gaussian Adaptive Resonance Theory Neural Network Classification Algorithm Applied to Supervised Land Cover Mapping Using Multitemporal Vegetation Index Data

Doug Muchoney and James Williamson

Abstract—Neural network classifiers have been shown to provide supervised classification results that significantly improve on traditional classification algorithms such as the Bayesian (maximum likelihood [ML]) classifier. While the predominant neural network architecture has been the feedforward multilayer perceptron known as backpropagation, Adaptive resonance theory (ART) neural networks offer advantages to the classification of optical remote sensing data for vegetation and land cover mapping. A significant advantage is that it does not require prior specification of the neural net structure, creating as many internal nodes as are needed to represent the calibration (training) data. The Gaussian ARTMAP classification algorithm bases the probability that input training samples belong to specific classes on the parameters of its Gaussian distributions: the means, standard deviations, and *a priori* probabilities. The performance of the Gaussian ARTMAP classification algorithm in terms of classification accuracy using independent validation data indicated was over 70% accurate when applied to an annual series of monthly 1-km advanced very high resolution radiometer (AVHRR) satellite normalized difference vegetation index (NDVI) data. The accuracies were comparable to those of fuzzy ARTMAP and a univariate decision tree, and significantly higher than a Bayesian classification algorithm. Algorithm testing is based on calibration and validation data developed and applied to Central America to map the International Geosphere-Biosphere Programme (IGBP) land cover classification system. Thus, it provides a realistic test of the algorithms for operational classification of a regional remote sensing and site dataset.

I. INTRODUCTION

A. MODIS 1-km Land Cover and Land Cover Change

BOSTON University, Boston, MA, will produce land cover and land cover change products globally at 1000 m and 0.25° resolutions on a quarter-annual basis using an annual temporal sequence of multispectral and multiresolution moderate resolution imaging spectroradiometer (MODIS) data. The primary purpose of MODIS land cover characterization [1] is to support global modeling indirectly by providing inputs to

MODIS algorithms that generate model parameters such as leaf area index/fraction of photosynthetically active radiation (LAI/FPAR), bidirectional distribution function (BRDF), and surface temperature, and directly by providing land cover inputs to models.

The principal inputs to the quarterly land cover products are monthly BRDF, BRDF-corrected nadir surface reflectances, vegetation index, snow cover, land surface temperature, and spatial texture for a one-year sequence [1]. We have examined several different types of decision tree (DT) and artificial neural network classification algorithms to develop the land cover product. The DT classifiers include univariate decision trees [2], multivariate decision trees [3], and hybrid trees [4]. The neural network algorithms include two Adaptive Resonance Theory (ART) systems: Fuzzy ARTMAP [5]–[7] and Gaussian ARTMAP [8], [9]. Evaluation of these ARTMAP and DT classification algorithms using several remote sensing data sets has shown that they produce comparable results that are consistently superior to those produced by maximum likelihood (ML) classification [10], [11]. The primary land cover classification system to be mapped is the 17-class IGBP classification [12].

B. Objectives

The purpose of this research is to compare the Gaussian ARTMAP classification algorithm with another ART artificial neural network, and with decision tree and traditional Bayesian (ML) classification algorithms. Algorithm testing is based on calibration and validation data developed and applied using an annual series of monthly 1-km advanced very high resolution radiometer (AVHRR) normalized difference vegetation index (NDVI) data to map multiple land cover and vegetation classification systems and parameters in Central America [13]. In this regard, it provides a real test of the algorithms for operational classification of a regional remote sensing and site dataset.

II. SUPERVISED CLASSIFICATION ALGORITHMS FOR LAND COVER MAPPING

A. Bayesian Classification Algorithm

The Bayesian or maximum-likelihood classification algorithm (MLC) calculates the probability that a pixel belongs to one of the set of possible classes based on the mean measurement vector for each class and the covariance matrix for each class by band [14]. The probability that a pixel belongs to a class

Manuscript received March 7, 2000; revised November 3, 2000. This work was supported by NASA Contract NAS5-31369, Proyecto Ambiental de Centroamerica (PROARCA-CAPAS), the Comisión de la Centroamericana de Ambiente y Desarrollo (CCAD), and The Nature Conservancy.

D. Muchoney was with the Center for Remote Sensing, Department of Geography, Boston University, Boston, MA 02215 USA. He is now with Conservation International, Center for Applied Biodiversity Science, Washington, DC 20036 USA (e-mail: d.muchoney@conservation.org).

J. Williamson is with the Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215 USA.

Publisher Item Identifier S 0196-2892(01)07627-6.

is based on the distance between the pixel value and a scaled and variance/covariance-corrected class mean [5]. The pixel is then assigned to that class to which the weighted distance is the lowest (i.e., to which it has the highest probability of belonging). The principal assumptions of MLC are that the input channels are Gaussian (normally) distributed and that the training data statistics for each class, which can be thought of as samples, are also normally distributed. The assumption of equal probabilities can be modified by applying prior probabilities [15], which neural networks can also use. [16]. MLC has a long history of use, and is an efficient algorithm that is easy to understand and modify. Weaknesses of MLC are that its assumptions of normality and of equal probability are often incorrect.

B. Decision Tree Classification Algorithm

Decision trees have been widely applied to remotely sensed data for land cover and vegetation characterization [17]–[20]. A decision tree is a divisive classification procedure that recursively partitions a data set into subdivisions based on tests defined at each branch (or node) in the tree [2]. A decision tree is composed of a root node comprising all of the data, a set of internal nodes (splits), and a set of terminal nodes (leaves). Each internal node in a decision tree has one parent node, and two or more descendant nodes. Using this framework, a data set is classified according to the decision surfaces defined by the tree, and class labels are assigned to each observation according to the leaf node into which the observation falls. The decision tree algorithms can use univariate [21], bivariate and multivariate [3] criteria [4]. Hybrid decision trees [10] use multiple classification algorithms on different subtrees within the framework of a single, larger decision tree structure.

C. Neural Network Classification Algorithms

Neural nets are complex and dense systems of nonlinear computational elements that are patterned after bioneurological systems that are composed of computational nodes that are linked by adaptive weights. Neural nets are self-organizing, adaptive and nonlinear [22]. For networks in general, gain control, vigilance and choice parameters determine how inputs are matched to existing categories of outputs and how “novelties” are detected. Neural nets must be both plastic, that is able to recognize new inputs, and stable, that is, insensitive to all but the important changes. Mathematically, the most important component is the transfer function which determines how a neuron will scale its response to incoming signals and produce an activation. The transfer functions are generally threshold-logic, hard-limit, continuous function (sigmoidal) nodes, and radial basis functions [23].

1) *Backpropagation*: The neural network backpropagation classification algorithm is composed of layers of neurons that are interconnected through weighted synapses. The first layer (F1) consists of the classification input variables and the layer output consists of a binary vector representing the output classes. Intermediate, hidden layers provide an internal representation of neural pathways through which input data are processed to arrive at output values or conclusions [1].

In a supervised mode, input variables are fed forward through the network to produce an output vector. During a following

backpropagation phase, the synapse weights are adjusted so that the network output vector more closely matches the desired output vector, which is a binary-coded representation of the training class. The network weights are adjusted by feeding the summed squared errors from the output layer back through the hidden layers to the input layer. The network cycles through the training set until the synapse weights have been adjusted so that the network output has converged, to an acceptable level, with the desired output. The trained neural network is then given new data and the internal synapses guide the processing flow through excitation and inhibition of neurons. This results in the assignment of the input data to the output classes [1]. Backpropagation requires that the number of internal layers be specified beforehand based on empirical results and understanding of the relationship between training/testing and the input image data features [24].

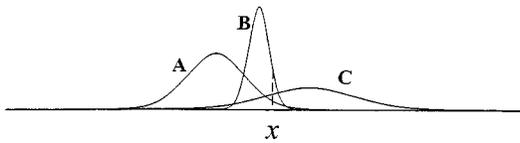
2) *Adaptive Resonance Theory*: ART networks are fundamentally different than backpropagation algorithms. The neural nets used in this analysis are from the class of ART [5] networks, including fuzzy ART [6], [7] and Gaussian ARTMAP [8], [9]. ART neural networks process inputs into categories, with the category formation being governed by a set of three parameters. The vigilance parameter (ρ) regulates how broad a category might be. The match function then determines if a selected category is sufficiently appropriate to meet the vigilance criteria. These categories are related to specific output classes in such a way that they represent multiple, nonlinear partitions of feature space that map to output classes [8].

ART models form stable recognition categories in response to arbitrary inputs. The original ART network categorized binary inputs using unsupervised or “teacher-less” learning. ART 2 is an extension of ART that allows for categorization of analog inputs. ART incrementally clusters inputs into stable categories, with the number of categories a function of the vigilance parameter ρ . The vigilance parameter determines how broad categories might be in feature space based on the networks distance metric. Thus, ART incrementally creates the number of categories that are needed to define clusters of input samples, with the inclusiveness of the categories inversely related to ρ [8].

The ART activation function provides an estimate of the likelihood that an input belongs to a specific category, while the match function determines if the category is similar enough to the input to satisfy the vigilance parameter. Activation of a category (F₂ node) can be interpreted as making a hypothesis about an input [8], [9].

ARTMAP: ARTMAP extends ART into a supervised classification mode by using the ARTs unsupervised clustering. ART categories (hidden units or F₂ nodes) learn predictions or mappings to output classes during training. If a chosen category makes the wrong prediction, the vigilance parameter is temporarily raised to the level needed to reset the category. If the chosen category satisfies the match function, the system resonates and learning has been achieved. This match tracking guarantees that for a given input, the category (or categories if distributed learning are used) that resonates has a better match than all other categories that are reset. So ARTMAP organizes the clustering of the data based on predictive feedback from the labels that it assigns to the categories (clusters), as well as by how the data are distributed in feature space [8].

Activation: Gaussian distributions with a priori probabilities.



Match: Gaussian distributions with unit height.

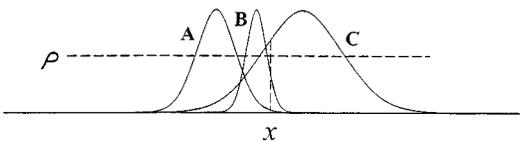


Fig. 1. Gaussian ARTMAP activation and match functions.

Fuzzy ARTMAP: Fuzzy ARTMAP, like ARTMAP, extends the binary ART 1 to the analog domain by using the \wedge AND fuzzy operator instead of the \cap logical intersection. Fuzzy ARTMAP input vectors are complement coded, which makes the denominator of the match criteria a constant. Each category (j) is initialized with a weight vector w_j . The choice function (α) selects the nonreset category with the highest activation that is determined by the size of the weight vector. If the choice parameter α is small, broad categories in feature space are favored. If α is large, categories with large weight vectors, and thus tight categories, are favored. The match criterion requires that a chosen category's weight vector be sufficiently close to the input vector [8]. ART classifiers have been found to be an improvement on traditional classifiers in some respects. ARTMAP dynamics are fast, stable, and scalable, overcoming common limitations of other neural networks such as backpropagation [1].

3) *Gaussian ARTMAP Classification Algorithm:* Gaussian ARTMAP is an adaptation to ART that is based on using Gaussian distributions to define the category choice functions. For each category, there is an associated *a priori* probability and a mean and variance in each input dimension. The Gaussian ART activation function evaluates the probability that an input belongs to a category's distribution, as well as the category's *a priori* probability (Fig. 1). The match function (Fig. 1) is based on how well the input fits the category's distribution, which is normalized to unit height. The likely class prediction is based on these activations [8], [9]. For Gaussian ARTMAP, high vigilance (ρ) means that more internal categories are created by the network to match input data to output categories; the categories are less broad in the feature space.

Gaussian ARTMAP accommodates choice and distributed learning. In choice learning, the maximally activated category is chosen. That is, the chosen category's match function satisfies the vigilance criterion, and the category resonates and learns the prediction. In distributed learning, each category is assigned credit based on the proportion of the net activation of *all* categories whose match function satisfies the vigilance criterion. When Gaussian ART is extended to Gaussian ARTMAP, the prediction of an output class during testing is akin to picking the class with the highest net probability. This is similar to methods of potential and radial basis functions. The activation of all categories sharing the same prediction are summed to yield the

most probable class prediction rather than basing the prediction solely on the maximum ART category as in Fuzzy ARTMAP.

Calibrating (training) Gaussian ARTMAP only requires estimating category means, covariances, and *a priori* probabilities from the training data. The Gaussian distributions themselves must be separable. Gaussian ARTMAP can fit data that vary between dimensions, but not that covary. This would require that each category store a covariance matrix. Like the MLC algorithm, Gaussian ARTMAP is based on the assumption that, for each category, the input channels are normally distributed. However, because multiple categories can map to a single output class, there is no assumption that the class conditional distributions are normal. Rather, the assumption is that they are mixtures of normal distributions [8], [9].

In a supervised mode, Gaussian ARTMAP chooses an output class with the highest conditional probability and also provides an indication of the strength of the relationship with the test data. Gaussian ARTMAP represents the input data density with separable Gaussian distributions, with the number and inclusivity of the distributions a function of the vigilance parameter. Both Fuzzy and Gaussian ARTMAP most efficiently represent data that are uncorrelated across dimensions.

The primary difference between Fuzzy ARTMAP and Gaussian ARTMAP is the use of different statistics to define each category. Rather than representing an interval within each dimension, as is the case with Fuzzy ARTMAP, Gaussian ARTMAP represents the mean and variance. In addition, Gaussian ARTMAP represents the *a priori* probabilities of the category. There is potential inefficiency in the fuzzy categories, which are defined by the minimum and maximum in each dimension in terms of their ability to support smooth generalizations in higher dimensions [8].

III. TEST SITE AND DATA

A. Site Data

The dimensions of the study area are 1169×1813 -km AVHRR pixels, bounded by 6° to 9° N and 77.22° to 93° W (Fig. 2). The regional site comprises southern Mexico, Guatemala, Belize, Honduras, El Salvador, Nicaragua, Costa Rica, and Panama. Central America includes a diverse array of natural and human-modified landscapes including broadleaf evergreen, deciduous and semi-deciduous forests, pine savanna and woodlands, swamp and mangrove forests, herbaceous wetlands, and agricultural types. As such, it provides an excellent test for regional classification.

The data for training and testing the supervised classifiers were developed by Boston University, the Stanford University Center for Conservation Biology, Stanford, CA, The Nature Conservancy, and the Central America Vegetation Working Group of the Comisión de la Centroamericana de Ambiente y Desarrollo (CCAD), which was established to support vegetation mapping and monitoring in Central America. Pairs of analysts delineated training/testing sites on Landsat thematic mapper (TM), Systeme pour l'Observation de la Terre Haute Resolution Visible (SPOT-HRV), and AVHRR data, and populated the database based on field, plot and ancillary map data at two workshops held in Guatemala and Nicaragua.

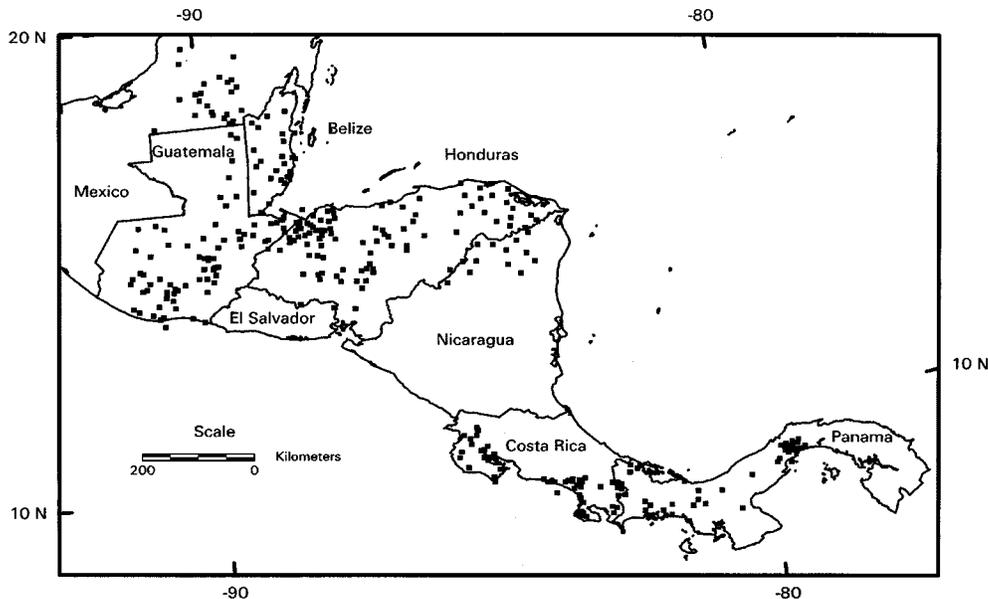


Fig. 2. Study area and test site locations.

428 sites were distributed among the 17 IGBP classes and extracted from 18 Landsat TM and 2 SPOT-XS multispectral (XS) scenes (Fig. 2, Table I). The criteria for selecting sites were that they were at least $2 \text{ km} \times 2 \text{ km}$ and that they be representative of the bioregions, vegetation, and land cover classes. The site polygons were also defined to be within larger patches of classes, with at least a 1-km buffer from the polygon boundary to the patch boundary. This was to ensure that misregistration and mislocation of the AVHRR data and the co-referenced TM or SPOT data would not permit a training polygon to actually represent land cover outside of the patch. After a quality assurance check was performed to ensure that the site data labels were correct, 20 sites were removed from the database because they were either obviously mislabeled or did not meet the minimum site size criteria resulting in 408 total sites.

Feature extraction and parameterizing the database involved assigning labels to appropriate categories of a suite of parameters based on the System for Terrestrial Ecosystem Parameterization (STEP) Model and database [25]. STEP provides for explicit description of the structural, functional, and compositional components of the vegetation and landscape tied to specific sites and plots. The primary purpose of STEP is to provide a comprehensive model of the land surface that can be used to train and test algorithms and to validate land surface products. Formal sites are established and described based on high-resolution remote sensing data, ancillary data, and field plot data. STEP can be used to translate multiple classification systems better than commonly used look-up table approaches. This accommodates the wide array of classifications used by various models to parameterize biophysical processes such as those of the biosphere-atmosphere transfer scheme (BATS) [26], Biome-BGC [27]–[31], land surface model (LSM) [32], the simple biosphere model (SiB) [33], and the SiB2 [34].

B. Remotely Sensed Data

The primary remotely sensed data used in this study are monthly composited AVHRR NDVI data for 1992–93 provided

TABLE I
SITE DISTRIBUTION BY IGBP LAND COVER CLASS

class	IGBP Class	sites	pixels (km^2)	mean area (km^2)
1	Evergreen needleleaf forest	42	1524	36.3
2	Evergreen broadleaf forest	131	3629	27.7
3	Deciduous needleleaf forest	0	0	0.0
4	Deciduous broadleaf forest	15	386	25.7
5	Mixed forest	25	847	33.9
6	Closed shrublands	4	156	39.0
7	Open shrublands	10	336	33.6
8	Woody savannas	12	469	39.1
9	Savannas	9	297	33.1
10	Grasslands	27	845	31.3
11	Permanent wetlands	40	842	21.1
12	Croplands	51	1968	38.6
13	Urban and built-up	16	269	16.8
14	Cropland mosaics	10	370	37.0
15	Snow/ice	0	0	
16	Barren or sparsely vegetated	3	32	10.8
17	Water bodies	13	721	55.5
	totals	408	12691	31.11

by the U.S. Geological Survey EROS Data Center (USGS-EDC), Boulder, CO [35], in lieu of MODIS data that will be used for generating global 1-km land cover when data become available following its recent launch on the Terra (EOS-AM) platform in 2000. NDVI is based on the relationship of reflected red and near-infrared reflectance that is highly correlated with both leaf area index (LAI) and biomass. The monthly data provide an indication of overall greenness of the vegetation, and can be used to characterize the phenology of vegetation. The AVHRR data were monthly composited using maximum NDVI to remove cloud and topographic effects and extreme off-nadir pixels [35], [36], as well as scan angle dependence of radiance [37]. Additional bidirectional reflectance distribution function (BRDF) or atmospheric correction was not employed.

TABLE II
REGIONAL IGBP CLASSIFICATION RESULTS (MEAN PERCENT CLASSIFICATION ACCURACY OF FIVE RUNS)

IGBP Class	Class Name	train/test pixels	DTC	MLC	Gaussian ARTMAP	Fuzzy ARTMAP
1	Evergreen needleleaf forest	1515	74.87	54.63	84.49	79.00
2	Evergreen broadleaf forest	3575	84.55	87.11	91.55	79.00
3	Deciduous needleleaf forest					
4	Deciduous broadleaf forest	370	63.82	63.26	75.41	83.90
5	Mixed forest	845	65.55	38.55	72.31	76.90
6	Closed shrublands	125	54.15	23.26	46.40	83.00
7	Open shrublands	335	65.66	37.71	82.69	93.00
8	Woody savannas	470	65.12	24.92	77.66	85.80
9	Savannas	60	34.25	5.36	15.00	78.10
10	Grasslands	845	70.48	49.26	78.22	81.20
11	Permanent wetlands	800	65.92	46.14	72.00	80.60
12	Croplands	1950	70.37	57.01	83.69	68.00
13	Urban and built-up	265	57.25	64.89	71.32	82.90
14	Cropland mosaics	365	68.62	35.10	76.99	91.70
15	Snow/ice					
16	Barren or sparsely vegetated	35	29.83	17.71	28.57	72.20
17	Water bodies	440	98.48	94.10	97.05	95.40
total/mean		11995	74.79	52.28	82.77	79.30

The use of the monthly-composited AVHRR data may be problematic [36]. Compositing is biased toward selecting off-nadir pixels, especially in forward-scanning views in winter months in the northern hemisphere [38]. As with any large-area projection, they also found that the effective mapping unit was geographically variable. In this case, errors due to the Goode's homolosine projection system and resampling methods [35] were most probably minimal compared to those caused by misregistration and varying viewing/illumination artifacts. Lack of sensor calibration confuses the temporal trajectory of the multitemporal NDVI signal [39]. A land/sea mask was applied to the AVHRR data as part of the processing. Temporal smoothing or generalization might enhance the meaning of the temporal signal [40], but was not applied to the data set used in this study. While the use of temporal metrics such as month of maximum NDVI and annual range of NDVI have been shown to improve classification accuracy [41], [42], the monthly data were used as simple features in the classification process.

The 1-km AVHRR-NDVI data lack the spectral resolution, radiometric calibration and locational accuracy qualities of MODIS data. NDVI is especially unsuited for mapping nonvegetated land-cover classes such as snow, ice, water, urban areas, and bare ground. The AVHRR data do, however, provide a regional multitemporal dataset, and maximum-value NDVI monthly compositing removed many of the atmospheric effects evident in the daily or ten-day composite data for the same period.

IV. EVALUATION CRITERIA, TESTING, AND RESULTS

A. Evaluation Criteria

The evaluation criteria for the classification algorithms are classification accuracy, training data requirements, and the algorithm's ability to generalize diverse training data to appropriate output classes. Given the cost of generating reliable site data for training, testing, and validation, the classification algorithm requirement for training data is an important consideration. These requirements include the number and quality of training sites.

TABLE III
GAUSSIAN ARTMAP ρ VALUES, CLASSIFICATION ACCURACY, AND F_2 NODES

ρ	accuracy (%)	F_2 nodes	% test samples not classified
0.001	91.20	427	0
0.01	91.20	427	0
0.1	91.20	427	0
0.2	91.20	428	0
0.3	91.32	434	0
0.4	91.00	448	0
0.5	92.02	502	0
0.6	92.88	635	0
0.7	93.89	880	0
0.8	94.85	1446	0
0.9	93.83	2833	0
0.95	91.10	4000	2.53
0.99	64.59	4000	11.03

Problems with site data include errors in labeling and representativeness of subpopulations.

MLC requires substantial training data, in the neighborhood of at least 10–30 times the number of training examples per class [43], [16]. For neural networks, there seems to be the promise of using smaller training sets [44]. Neural networks in general have been found to be robust to training site heterogeneity [24] and missing data. Backpropagation can require extensive training, while ART is designed to be stable enough to preserve significant past learning while still allowing new information to be incorporated in the neural network structure as such information appears in the data input stream [1]. Decision trees are tolerant of noisy training data and have been used to remove noisy/dirty training data [10].

B. Testing the Supervised Classification Algorithms

From the 408 sites, five sets of calibration (80%), and the validation (20%) pixel were randomly generated by class (strat-

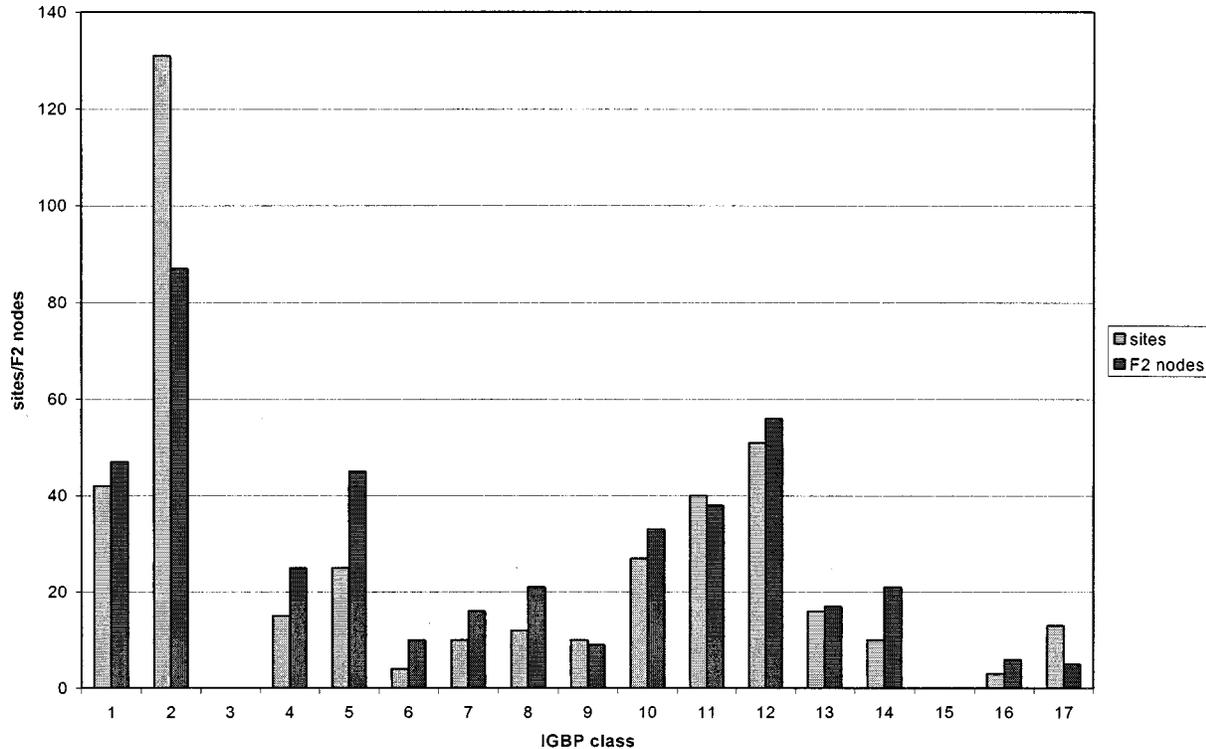


Fig. 3. Relationship of number of sites to F₂ nodes by class.

ified random sampling) to allow for independent training and accuracy assessment. In this approach, the algorithm calibrates (trains) on 80% of the pixels in each class, and its accuracy is assessed (validated) for the remaining 20% of pixels, which are unseen. Since results are sometimes dependent on the actual random selection of pixels, we repeated the selection procedure five times, thus providing five sets of training and testing pixels to the classification and accuracy assessment process.

Supervised classification based on the five random pixel splits was performed using each classification algorithm applied to the 17-class IGBP land cover system by training and testing on the monthly-composited NOAA AVHRR NDVI data for 12 months of 1992 and 1993 [35]. The IGBP labels for the five calibration and validation splits of the 408 sites were applied to the Gaussian ARTMAP, Fuzzy ARTMAP, decision tree, and ML (Bayesian) classification algorithms. For all approaches, the error rates were similar for all five of the calibration-validation pixel splits.

V. RESULTS OF PIXEL CALIBRATION AND VALIDATION

A. Bayesian Classification Algorithm

The mean overall accuracy of the MLC algorithm as applied without prior probabilities with one statistic per class was poor in comparison to the other algorithms, with a mean accuracy of 52.58 percent and a range of approximately 49–53%. A significant problem with the MLC algorithm is singularities (zeros) that occur in the covariance matrix and result in division by zero errors.

B. Decision Tree Classification Algorithm

The decision tree classification algorithm tested was a univariate, the results of which are provided in Table II. The mean overall accuracy of the five runs was 74.79%.

TABLE IV
GAUSSIAN ARTMAP ERROR RATES BY CLASSIFICATION RUN

IGBP class	classification run ($\rho=0.6; g=30$)					test pixels	mean accuracy
	1	2	3	4	5		
1	84.16	85.15	82.51	82.51	88.12	303	84.49
2	92.45	91.89	92.59	90.07	90.77	715	91.55
3	0	0	0	0	0	0	0
4	72.97	77.03	81.08	71.62	74.32	74	75.41
5	70.41	71.01	73.37	70.41	76.33	169	72.31
6	48	44	52	48	40	25	46.40
7	85.07	76.12	94.03	77.61	80.6	67	82.69
8	75.53	81.91	76.6	79.79	74.47	94	77.66
9	25	16.67	8.33	25	0	12	15.00
10	71.6	85.21	75.15	78.7	80.47	169	78.22
11	66.27	68.05	69.23	68.64	68.64	160	72.00
12	86.41	82.05	84.36	84.87	80.77	390	83.69
13	69.81	64.15	75.47	79.25	67.92	53	71.32
14	79.45	78.08	75.34	73.97	78.08	73	76.99
15	0	0	0	0	0	0	0
16	28.57	14.29	14.29	42.86	42.86	7	28.57
17	97.73	92.05	98.86	97.73	98.86	88	97.05
overall	82.74	82.74	83.41	82.24	82.70	2399	82.77

C. Fuzzy ARTMAP Classification Algorithm

For the Fuzzy ARTMAP classification, the choice parameter α was applied through its range of 0.01–0.99 to determine the optimum value based on the classification accuracy of unseen test pixels. An α value of 0.95 was selected, and applied to the five train/test data splits. The mean overall accuracy for the Fuzzy ARTMAP runs is 79.30% (Table II).

D. Gaussian ARTMAP Classification Algorithm

For the Gaussian ARTMAP algorithm, the vigilance parameter ρ was applied through its range of 0.01–0.99 to determine the optimum value based on the classification accuracy of unseen test pixels. As expected, a smaller vigilance resulted in fewer F₂ nodes being created to represent the data and as ρ increased, the number of F₂ categories also increased. The rela-

TABLE V
 GAUSSIAN ARTMAP IGBP CLASSIFICATION ACCURACY CONTINGENCY TABLE

IGBP Class	Reference Data IGBP Class (percent 1 km ² pixel agreement)																	test pixels
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	65.35	11.88			2.64		0.66			5.94	6.27	6.93	0.33					303
2	5.03	67.13		0.42	1.40			0.42		0.84	0.70	2.66		1.12				715
3																		0
4	6.76	16.22		50.00	1.35					5.41		17.57		1.35			1.35	74
5	15.38	6.51		1.78	49.70		0.59	0.59		9.47	5.33	10.65						169
6	4.00	24.00				4.00			4.00	24.00		36.00		4.00				25
7	10.45	1.49					44.78	11.94		13.43		16.42					1.49	67
8	10.64	13.83		1.06			1.06	38.30		7.45	5.32	21.28	1.06					94
9	8.33	33.33							0.00	8.33	16.67	25.00		8.33				12
10	1.18	2.95		1.78	1.78		1.18			78.11	0.59	12.43						169
11	15.00	14.38		0.63	1.88		0.63	4.38		3.13	49.38	6.88	0.63	0.63			2.50	160
12	8.72	6.21		0.26	1.20					8.46	1.25	71.28		0.51				390
13	5.66	5.66			1.89			3.77			7.55	11.32	64.15					53
14		38.36		2.74						10.96		6.85	2.74	38.36				73
15																		0
16		42.86					14.29			42.86						0.00		7
17													2.25	1.12			96.63	69
total																		2400

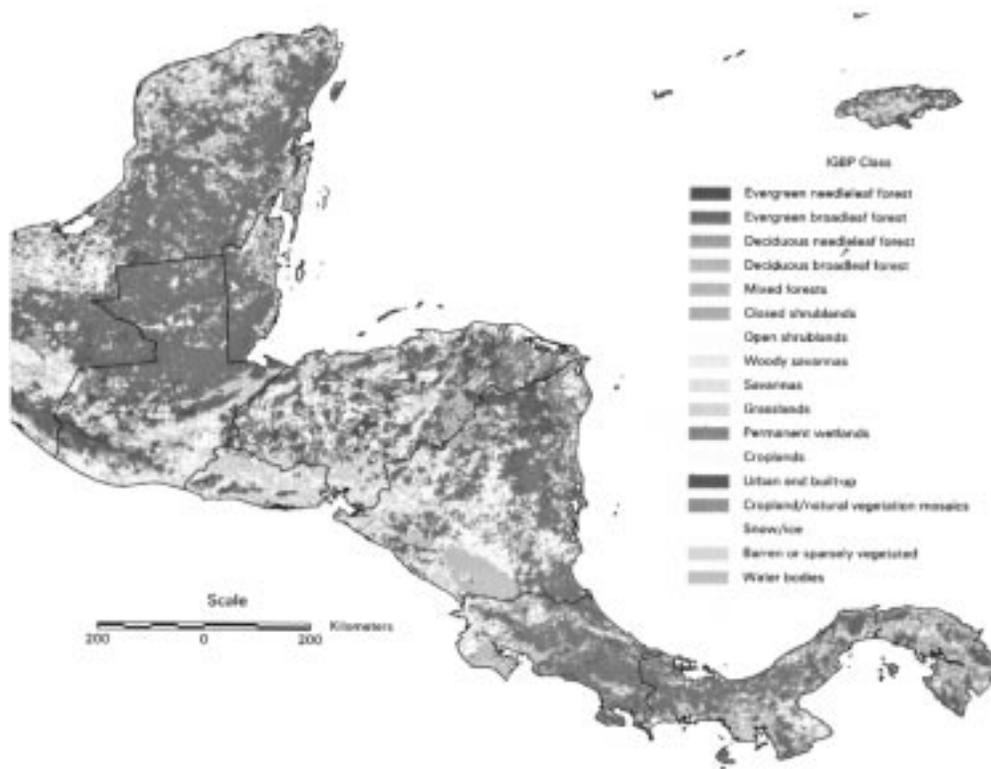


Fig. 4. Gaussian ARTMAP IGBP land cover classification.

tionship of ρ , F_2 categories, and per-class accuracy is presented in Table III. Accuracy and the number of F_2 nodes increases with accuracy up to $\rho = 0.8$, at which time, accuracy begins to decrease. More F_2 nodes are being created to (over) fit the data, to the point where validation (test) data are no longer being classified, that is, there is no longer any category that meets the match criteria.

Based on this analysis, parameter values of $\rho = 0.6$ and a variance of $\gamma = 30$ were selected to train each of the five iterations based on exploring the relationship of the vigilance parameter to training. This vigilance value is a compromise between using a high vigilance that tends to fit the training data

(exemplars) well to the output classes and the need to generalize training data to the output classes. Fig. 3 presents the relationship of the number of sites to the number of F_2 nodes that were created for each class. Some classes require fewer nodes than sites, while others require more. Gaussian ARTMAP also generates per-class probability estimates, indicating the probability that an input fits a class output. It is also possible to map individual F_2 categories rather than their coinciding class label. Intersecting the mapping of F_2 categories with site distribution confirmed that some sites required multiple F_2 categories to represent them, while other sites generalized well and could be represented by a single F_2 node.

The error rates for the Gaussian ARTMAP runs are reported in Table IV. Both per-class and overall accuracy were relatively stable for each of the five calibration/validation iterations (range = 80.92–83.86%, $SD = 0.945$), and the mean overall accuracy for the IGBP Gaussian ARTMAP classifications is 82.23%. The example contingency table of run 1 (Table V) presents the per-class errors and interclass error structure associated with Gaussian ARTMAP. The mean accuracy of the five Gaussian ARTMAP iterations is 82.77%. A regional map was produced using all calibration and validation data using Gaussian ARTMAP (Fig. 4).

VI. CONCLUSIONS

As Table II indicates, the primary forest classes (Classes 1 and 2) were mapped well by all algorithms, while Closed Shrubland and Savanna (Classes 6 and 9) were particularly problematic for the DT and Gaussian ARTMAP algorithms. Testing of the Gaussian ARTMAP classification algorithm can be summarized by the following.

- 1) It produced classification accuracies that are better than univariate decision tree and Fuzzy ARTMAP classification algorithms. Applying the IGBP classification system, the Gaussian ARTMAP averaged almost 83%, and Fuzzy ART neural network, 79%, and the DT almost 75%. Gaussian ARTMAP does not require prior specification of the network structure. It is flexible in that it generates a sufficient number of internal categories that are needed to match inputs to outputs. Its parameters can be varied to develop internal categories that can either generalize or designed to fit the input data specifically.
- 2) Predictions are based on the mean, variance, and the *a priori* probabilities of the category, making interpretation of the Gaussian ARTMAP network rather straightforward.
- 3) The use of pixel-based splits for calibration and validation is problematic for some algorithms. When using pixel-based sampling to calibrate and validate site-data subsets, the algorithm is presented with one or more examples (pixels) from each site. This has been shown to provide higher agreement indices than when polygon splitting is performed [41], [45].
- 4) The number of F_2 categories varies for land cover types indicating that some classes generalize better than others.
- 5) The probability estimate of Gaussian ARTMAP is useful for evaluating the relative strength of the prediction.
- 6) The individual F_2 nodes can be mapped, and their relationship to individual sites can be evaluated. This provides a useful indicator of the number of F_2 nodes that are required to represent a specific site. If a site requires multiple nodes to represent or capture its variability, then it most probably is a heterogeneous site. It is possible to determine what site or sites that are represented by one or more nodes. This feature can be used to indicate potentially mislabeled or mixed sites.

As such, Gaussian ARTMAP has been demonstrated to be a viable algorithm for supervised classification of remotely sensed data.

ACKNOWLEDGMENT

The authors would like to thank J. Correau of PROARCA-CAPAS, the members of the Central America Vegetation Working Group, and R. Sayre and X. Li of The Nature Conservancy for their contribution to this research. The authors would also like to acknowledge the support of J. Borak, H. Chi, M. Friedl, J. Hodges, S. Gopal, D. McIver, and A. Strahler of the Boston University Center for Remote Sensing, Department of Geography, Boston, MA.

REFERENCES

- [1] A. Strahler, D. Muchoney, J. Borak, G. Feng, M. Friedl, S. Gopal, J. Hodges, E. Lambin, D. McIver, A. Moody, C. Schaaf, and C. Woodcock, "MODIS land cover product algorithm theoretical basis document (ATBD) version 5.0.," Boston Univ., Boston, MA, 1999.
- [2] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
- [3] C. E. Brodley and P. E. Utgoff, "Multivariate decision trees," *Mach. Learn.*, vol. 19, pp. 45–77, 1995.
- [4] C. E. Brodley, "Recursive automatic bias selection for classifier construction," *Mach. Learn.*, vol. 20, pp. 63–94, 1995.
- [5] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, vol. 4, pp. 566–588, 1991.
- [6] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.
- [7] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ART: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, vol. 3, pp. 698–713, 1992.
- [8] J. R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Networks*, vol. 9, pp. 881–897, 1996.
- [9] —, "A constructive, incremental-learning network for mixture modeling and classification," *Neural Comput.*, vol. 9, pp. 1517–1543, 1996.
- [10] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61, pp. 399–409, 1998.
- [11] S. Gopal, C. Woodcock, and A. H. Strahler, "Fuzzy neural network classification of global land cover from a 1° AVHRR data set," *Remote Sens. Environ.*, vol. 67, no. 2, pp. 230–243, 1999.
- [12] T. R. Loveland and A. Belward, "The IGBP-DIS global 1 km land cover data set, DISCover first results," *Int. J. Remote Sensing*, vol. 18, no. 15, pp. 3289–3295, 1997.
- [13] D. M. Muchoney, J. Borak, H. Chi, M. Friedl, J. Hodges, N. Morrow, and A. Strahler, "Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America," *Int. J. Remote Sensing*, vol. 21, pp. 1115–1138, 2000.
- [14] J. R. Jensen, *Introductory Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1986, p. 379.
- [15] A. H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sens. Environ.*, vol. 10, pp. 135–163, 1980.
- [16] G. M. Foody, "Using prior knowledge in artificial neural network classification with a minimal training set," *Int. J. Remote Sensing*, vol. 16, no. 2, pp. 799–810, 1995.
- [17] D. E. Brown, V. Corruble, and C. L. Pittard, "A comparison of decision tree classifiers with backpropagation neural networks for multimodal classification problems," *Pattern Recognit.*, vol. 26, pp. 953–961, 1993.
- [18] R. C. Dubayah, J. Michaelsen, and D. S. Schimel, "Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys," *J. Vegetation Sci.*, vol. 5, pp. 673–686, 1994.
- [19] J. Michaelsen, D. S. Schimel, M. A. Friedl, F. W. Davis, and R. C. Dubayah, "Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys," *J. Vegetation Sci.*, vol. 5, pp. 673–686, 1994.
- [20] M. Hansen, R. Dubayah, and R. DeFries, "Classification trees: An alternative to traditional land cover classifiers," *Int. J. Remote Sensing*, vol. 17, pp. 1075–1081, 1996.

- [21] L. Breiman, J. H. Friedman, R. A. Olsen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
- [22] R. P. Lippmann, "Pattern classification using neural networks," *IEEE Commun. Mag.*, vol. 27, pp. 47–55, 1989.
- [23] C. T. Hairston, "The neurological basis for neural computations," in *Handbook of Neural Computing Applications*, C. Parten, C. Hairston, and R. Pap, Eds. San Diego, CA: Academic, 1990, pp. 129–183.
- [24] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 981–996, July 1995.
- [25] D. M. Muchoney, A. Strahler, J. Hodges, and J. Locastro, "The IGBP DISCover confidence sites and the system for terrestrial ecosystem parameterization: Tools for validating global land cover data," *Photogramm. Eng. Remote Sensing*, vol. 65, no. 9, pp. 1061–1067, 1999.
- [26] R. E. Dickinson, A. Henderson-Sellers, and P. J. Kennedy, "Biosphere atmosphere transfer scheme (BATS) version 1 as coupled to the NCAR community climate model," Nat. Center Atmos. Res., Tech. Note, NCCR, Boulder, CO, 1993.
- [27] R. B. Myneni, R. Ramakrishna, and S. W. Running, "Estimation of global leaf area index and absorbed PAR using radiative transfer models," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 1380–1393, 1997.
- [28] R. Nemani and S. W. Running, "Implementing a hierarchical global vegetation classification in ecosystem function models," *J. Vegetation Sci.*, vol. 7, pp. 337–346, 1996.
- [29] ———, "Land cover characterization using multitemporal red, near-IR, and thermal-IR data from NOAA/AVHRR," *Ecol. Applicat.*, vol. 7, pp. 79–90, 1997.
- [30] S. W. Running, T. R. Loveland, and L. L. Pierce, "A vegetation classification logic based on remote sensing for use in global biogeochemical models," *Ambio*, vol. 23, pp. 77–81, 1994.
- [31] S. W. Running, T. R. Loveland, L. L. Pierce, R. Nemani, and E. R. Hunt, Jr., "A remote sensing based vegetation classification logic for global land cover analysis," *Remote Sens. Environ.*, vol. 51, pp. 39–48, 1995.
- [32] G. B. Bonan, "A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: technical description and user's guide," NCAR/TN-417+STR, NCAR Tech. Note, Nat. Center Atmos. Res., Boulder, CO, 1996.
- [33] P. J. Sellers, Y. Mintz, Y. C. Sud, and A. Dalcher, "A simple biosphere model (SiB) for use within general circulation models," *J. Atmos. Sci.*, vol. 43, pp. 503–531, 1986.
- [34] P. J. Sellers, D. A. Randall, and G. J. Collatz *et al.*, "A revised land surface parameterization (SiB2) for atmospheric GCMs—Part 1: Model formulation," *J. Climate*, vol. 9, pp. 676–705, 1996.
- [35] J. C. Eidsensink and J. L. Faundeen, "The 1-km AVHRR global land data set: First stages in implementation," *Int. J. Remote Sensing*, vol. 15, pp. 3443–3462, 1994.
- [36] B. Holben, "Characteristics of maximum-value composite images from multitemporal AVHRR data," *Int. J. Remote Sensing*, vol. 7, pp. 1417–1434, 1986.
- [37] M. J. Duggin, D. Piwinski, V. Whitehead, and G. Ryland, "Scan-angle dependence of radiance recorded by the NOAA-AVHRR," *Proc. SPIE*, vol. 363, pp. 98–101, 1982.
- [38] Z.-L. Zhu and L. Yang, "Characteristics of the 1-km AVHRR data set for North America," *Int. J. Remote Sensing*, vol. 17, pp. 1915–1924, 1996.
- [39] J. Cihlar, "Identification of contaminated pixels in AVHRR composite images for studies of land biosphere," *Remote Sens. Environ.*, vol. 56, pp. 149–163, 1996.
- [40] A. Van Dijk, S. I. Calhs, C. M. Sakamoto, and W. L. Decker, "Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data," *Photogramm. Eng. Remote Sensing*, vol. 53, pp. 1059–1067, 1987.
- [41] D. M. Muchoney and A. Strahler, "Pixel and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data," *Remote Sens. Environ.*, to be published.
- [42] R. DeFries, M. Hansen, and J. Townshend, "Global discrimination of land cover types from metrics derived from AVHRR pathfinder data," *Remote Sens. Environ.*, vol. 56, pp. 209–222, 1995.
- [43] J. Piper, "Variability and bias in experimentally measured classifier error rates," *Pattern Recognit. Lett.*, vol. 13, pp. 685–692, 1992.
- [44] G. F. Hepner, T. Logan, N. Ritter, and N. Bryant, "Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification," *Photogramm. Eng. Remote Sensing*, vol. 56, no. 4, pp. 469–471, 1990.
- [45] M. Friedl, C. Woodcock, S. Gopal, D. Muchoney, A. Strahler, and C. Barker-Schaaf, "A note on procedures for accuracy assessment in land cover maps derived from AVHRR data," *Int. J. Remote Sensing*, to be published.

Doug Muchoney received the B.S. degree in forestry from Virginia Tech, Blacksburg, the M.S. degree in geographical and cartographical science from George Mason University, Fairfax, VA, and the Ph.D. degree in geography from Boston University, Boston, MA, in 1982 and 1991, respectively.

He is currently an Ecologist/Remote Sensing Scientist with the Center for Applied Biodiversity Science, Conservation International, Washington, DC.

James Williamson received the B.S. degree in psychology from the University of Massachusetts, Amherst, in 1987, and the Ph.D. degree in cognitive and neural systems from Boston University, Boston, MA, in 1995.

He is currently a Research Assistant Professor with the Department of Cognitive and Neural Systems, Boston University. His specialization is image processing and object recognition, including self-organization and topographic maps, and perceptual grouping.