# ART Neural Networks for Remote Sensing Image Analysis

Gail A. Carpenter, Marin N. Gjaja, Sucharita Gopal, Curtis E. Woodcock

Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
677 Beacon Street
Boston University
Boston, MA 02215

*ICIAP'99*: 10th International Conference on Image Analysis and Processing
September 27-29, 1999
Venice, Italy

Technical Report CAS/CNS TR-99-007
Boston, MA: Boston University

**Topic areas** –

Neural networks and perceptual learning
Pattern classification
Remote sensing and GIS

**Abstract** –

ART and ARTMAP neural networks for adaptive recognition and prediction have been applied to a variety of problems, including automatic mapping from remote sensing satellite measurements, parts design retrieval at the Boeing Company, medical database prediction, and robot vision. This paper features a self-contained introduction to ART and ARTMAP dynamics. An application of these networks to image processing is illustrated by means of a remote sensing example. The basic ART and ARTMAP networks feature winner-take-all (WTA) competitive coding, which groups inputs into discrete recognition categories. WTA coding in these networks enables fast learning, which allows the network to encode important rare cases but which may lead to inefficient category proliferation with noisy training inputs. This problem is partially solved by ART-EMAP, which use WTA coding for learning but distributed category representations for test-set prediction. Recently developed ART models (dART and dARTMAP) retain stable coding, recognition, and prediction, but allow arbitrarily distributed category representation during learning as well as performance.

## 1 ART and ARTMAP Neural Networks

Adaptive resonance theory originated from an analysis of human cognitive information processing and stable coding in a complex input environment (Grossberg, 1976, 1980). An evolving series of ART neural network models have added new principles to the early theory and have realized these principles as quantitative systems that can be applied to problems of category learning, recognition, and prediction. Each ART network forms stable recognition categories in response to arbitrary input sequences with either fast or slow learning regimes (Section 2). The first ART model, ART 1 (Carpenter and Grossberg, 1987a), was an unsupervised learning system to categorize binary input patterns. ART 2 (Carpenter and Grossberg, 1987b) and fuzzy ART (Carpenter, Grossberg, and Rosen, 1991) extend the ART 1 domain to categorize analog as well as binary input patterns.

Supervised ART architectures, called ARTMAP systems, self-organize arbitrary mappings from input vectors, representing features such as spectral values and terrain variables, to output vectors, representing predictions such as vegetation classes in a remote sensing application (Section 3). Internal ARTMAP control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error in an on-line setting. Binary ART 1 computations are the foundation of the first ARTMAP network (Carpenter, Grossberg, and Reynolds, 1991), which therefore learns binary maps. When fuzzy ART replaces ART 1 in an ARTMAP system, the resulting fuzzy ARTMAP architecture (Carpenter et al., 1992) rapidly learns stable mappings between analog or binary input and output vectors.

Recently fuzzy ARTMAP has become the basis of new methodologies for producing maps from satellite data (Carpenter et al., 1997a, 1997b, 1998; Gopal, Sklarew, & Lambin, 1994). A simplified version of this problem (Section 4) introduces and illustrates the dynamics of fuzzy ARTMAP networks. Other applications of unsupervised ART networks and supervised ARTMAP networks include a Boeing parts design retrieval system (Caudell et al. , 1994), robot sensory-motor control (Bachelder, Waxman, & Seibert, 1993; Baloch & Waxman, 1991; Dubrawski & Crowley, 1994a), machine vision (Caudell & Healy, 1994), 3D object recognition (Seibert & Waxman, 1992), Macintosh operating system software (Johnson, 1993), robot navigation (Dubrawski & Crowley, 1994b), automatic target recognition (Bernardon & Carrick, 1995; Koch et al., 1995; Waxman et al., 1995), electrocardiogram wave recognition (Ham & Han, 1993; Suzuki, Abe, & Ono, 1993), prediction of protein secondary structure (Mehta, Vij, & Rabelo, 1993), air quality monitoring (Wienke, Xie, & Hopke, 1994), strength prediction for concrete mixes (Kasperkiewicz, Racz, & Dubrawski, 1994), signature verification (Murshed, Bortolozzi, & Sabourin, 1995), tool failure monitoring (Ly & Choi, 1994; Tarng, Li, & Chen, 1994), chemical analysis from UV and IR spectra (Wienke & Kateman, 1994), frequency selective surface design for electromagnetic system devices (Christodoulou et al., 1995), face recognition (Seibert & Waxman, 1993), Chinese character recognition (Gan & Lua, 1992), and analysis of musical scores (Gjerdingen, 1990).

## 2 ART Dynamics

The central feature of all ART systems is a pattern matching process that compares the current input with a learned category representation, or active hypothesis, selected by the input. This matching process leads either to a resonant state which focuses attention and triggers category learning or to a self-regulating parallel memory search which always leads to a resonant sate, unless the network's memory capacity is exceeded. If the search ends with selection of an established category, then the category's learned representation may be refined to incorporate

new information from the current input. If the search ends by selecting a previously untrained node, the ART network establishes a new category.

Figure 1 illustrates the ART search cycle. During ART search, an input vector **A** registers itself as a pattern **x** of activity across level $F_1$ (Figure 1a). Converging and diverging $F_1 \rightarrow F_2$ adaptive filter pathways, each weighted by a long term memory (LTM) trace, or adaptive weight, transform **x** into a net input vector **T** to level $F_2$. The internal competitive dynamics of $F_2$ contrast-enhance vector **T**, generating a compressed activity vector **y** across $F_2$. In ART 1 and fuzzy ART, strong competition selects the $F_2$ node that receives the maximal $F_1 \rightarrow F_2$ input

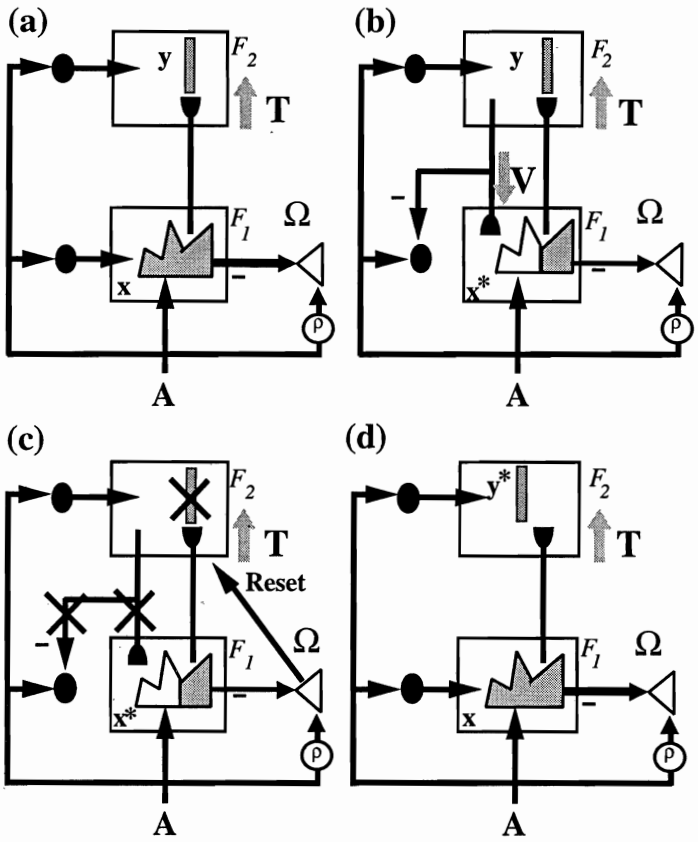


**Figure 1**. ART search for an $F_2$ code. (a) The input vector **A** generates the $F_1$ activity vector **x** as it activates the orienting subsystem Ω. Activity **x** both inhibits Ω and generates an $F_1 \rightarrow F_2$ signal. A bottom-up adaptive filter transforms **x** into the $F_2$ input vector **T**, which activates the STM pattern **y** across $F_2$. (b) A top-down adaptive filter transforms **y** into the category representation vector **V**. Where **V** mismatches **A**, $F_1$ registers a diminished STM activity pattern $\mathbf{x}^*$. The resulting reduction of total STM reduces the total inhibitory signal from $F_1$ to Ω. (c) If the ART matching criterion fails, Ω releases a nonspecific signal that resets the STM pattern **y** at $F_2$. (d) Since reset inhibits **y**, it also eliminates the top-down signal **V**, so **x** can be reinstated at $F_1$. However, enduring traces of the prior reset allow **x** to activate a different STM pattern $\mathbf{y}^*$ at $F_2$. If the top-down signal due to $\mathbf{y}^*$ also mismatches **A** at $F_1$, then the search for an $F_2$ code that satisfies the matching criterion continues. (Carpenter & Grossberg, 1987a)

component $T_J$. Only one component $(y_J)$ of **y** remains positive after this choice takes place. Activation of such a winner-take-all node selects category $J$ for the input pattern **A**.

Activation of an $F_2$ node may be interpreted as "making a hypothesis" about an input **A**. After sending the $F_2$ activity vector **y** through top-down adaptive pathways, a filtered vector **V** becomes the $F_2 \rightarrow F_1$ input (Figure 1b). The ART network matches the "expectation" pattern **V** of the active category against the current input pattern, or exemplar, **A**. This matching process typically changes the $F_1$ activity pattern **x**, suppressing activation of all features in **A** that are not confirmed by **V**. The resultant pattern $\mathbf{x}^*$ represents the features to which the network "pays attention." If the expectation **V** is close enough to the input **A**, then a state of resonance occurs, with the matched pattern $\mathbf{x}^*$ defining an attentional focus. The resonant state persists long enough for weight adaptation to occur; hence the term *adaptive resonance* theory. The fact that ART networks encode only attended features $\mathbf{x}^*$ rather than all input features **A** is directly responsible for ART code stability.

A dimensionless parameter called *vigilance* defines the criterion of an acceptable match. Vigilance specifies what fraction of the bottom-up input **A** must remain in the matched $F_1$ pattern $\mathbf{x}^*$ in order for resonance to occur. In ARTMAP, vigilance becomes an internally controlled variable, rather than the fixed parameter of ART. Because vigilance then varies across learning trials, a single ARTMAP system can encode widely differing degrees of generalization, or code compression. Low vigilance allows broad generalization, coarse categories, and abstract representations. High vigilance leads to narrow generalization, fine categories, and specific representations. At the very high vigilance limit, category learning reduces to exemplar learning. Varying vigilance levels allow a single ART system to recognize both abstract categories, such as faces and dogs, and individual faces and dogs.

ART memory search, or hypothesis testing, begins when the top-down expectation **V** determines that the bottom-up input **A** is too novel, or unexpected, with respect to the chosen category to satisfy the vigilance criterion. Search leads to selection of a better recognition code to represent input **A** at level $F_2$. An *orienting subsystem* $\Omega$ controls the search process. The orienting subsystem interacts with the attentional subsystem, as in Figures 1b and 1c, to enable the network to learn about novel inputs without risking unselective forgetting of its previous knowledge. ART 3 (Carpenter & Grossberg, 1990) implements parallel distributed search as a medium-term memory (MTM) process, as needed for distributed recognition codes.

ART search prevents associations from forming between **y** and $\mathbf{x}^*$ if $\mathbf{x}^*$ is too different from **A** to satisfy the vigilance criterion. The search process resets **y** before such an association can form. If the search ends upon a familiar category, then that category's representation may be refined in light of new information carried by **A**. If the search ends upon an uncommitted $F_2$ node, then **A** begins a new category. An ART *choice parameter* controls how deeply the search proceeds before selecting an uncommitted node. As learning self-stabilizes, all inputs coded by a category access it directly and search is automatically disengaged.

## 3 ARTMAP

ARTMAP networks for supervised learning self-organize mappings from input vectors, representing features such as spectral band values and terrain variables of a pixel, to output vectors, representing predictions such as the vegetation class of the site in which the pixel is

located. The original binary ARTMAP (Carpenter, Grossberg, & Reynolds, 1991) incorporates two ART 1 modules, $\mathrm{ART}_a$ and $\mathrm{ART}_b$, which are linked by a *map field* $F^{ab}$ (Figure 2). During supervised learning, $\mathrm{ART}_a$ receives a stream of patterns $\left\{\mathbf{a}^{(n)}\right\}$ and $\mathrm{ART}_b$ receives a stream of patterns $\left\{\mathbf{b}^{(n)}\right\}$, where $\mathbf{b}^{(n)}$ is the correct prediction given $\mathbf{a}^{(n)}$. An associative learning network and an internal controller link these modules to make the ARTMAP system operate in real time. The controller creates the minimal number of $\mathrm{ART}_a$ recognition categories, or "hidden units," needed to meet accuracy criteria. A minimax learning rule enables ARTMAP to learn quickly, efficiently, and accurately as it conjointly minimizes predictive error and maximizes code compression. This scheme automatically links predictive success to category size on a trial-by-trial basis using only local operations. It works by increasing the $\mathrm{ART}_a$ vigilance parameter $\rho_a$ by the minimal amount needed to correct a predictive error at $\mathrm{ART}_b$.

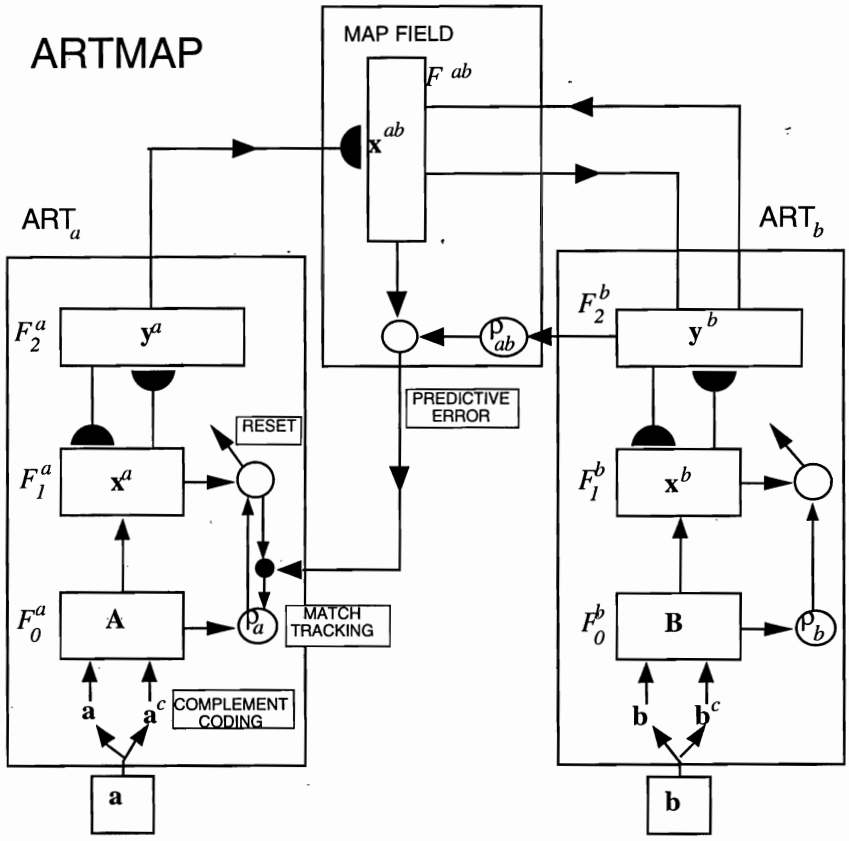


**Figure 2**. ARTMAP architecture. The $\mathrm{ART}_a$ complement coding preprocessor transforms the $M_a$-vector $\mathbf{a}$ into the $2M_a$-vector $\mathbf{A}=\left(\mathbf{a},\mathbf{a}^c\right)$ at the $\mathrm{ART}_a$ field $F_0^a$. $\mathbf{A}$ is the input vector to the $\mathrm{ART}_a$ field $F_1^a$. Similarly, the input to $F_1^b$ is the $2M_b$-vector $\mathbf{B}=\left(\mathbf{b},\mathbf{b}^c\right)$. When $ART_b$ disconfirms a prediction of $\mathrm{ART}_a$, map field inhibition induces the match tracking process. Match tracking raises the $\mathrm{ART}_a$ vigilance $\rho_a$ to just above the $F_1^a$-to-$F_0^a$ match ratio $\left|\mathbf{x}^a\right|/|\mathbf{A}|$. This triggers an $\mathrm{ART}_a$ search which leads either to an $\mathrm{ART}_a$ category that correctly predicts $\mathbf{b}$ or to a previously uncommitted $\mathrm{ART}_a$ category node. (Carpenter, Grossberg, & Reynolds, 1991)

At the map field an ARTMAP network forms associations between categories via outstar learning and triggers search, via a *match tracking* rule, when a training set input fails to make a correct prediction. Match tracking increases the $\mathrm{ART}_a$ vigilance parameter $\rho_a$ in response to a predictive error at $\mathrm{ART}_b$. A *baseline vigilance* parameter $\overline{\rho}_a$ calibrates a minimum confidence level at which $\mathrm{ART}_a$ will accept a chosen category. Lower values of $\overline{\rho}_a$ allow larger categories to form, maximizing code compression. Initially, $\rho_a = \overline{\rho}_a$. During training, a predictive failure at $\mathrm{ART}_b$ increases $\rho_a$ just enough to trigger an $\mathrm{ART}_a$ search. Match tracking sacrifices the minimum amount of compression necessary to correct the predictive error. Hypothesis testing selects a new ART category, which focuses attention on a cluster of $\mathbf{a}^{(n)}$ input features that is better able to predict $\mathbf{b}^{(n)}$. With fast learning, match tracking allows a single ARTMAP system to learn a different prediction for a rare event than for a cloud of similar frequent events in which it is embedded. Fuzzy ARTMAP (Carpenter et al., 1992) substitutes fuzzy ART for ART 1.

## 4 An ARTMAP Prototype Application: Satellite Remote Sensing

Mapping vegetation from satellite remote sensing data has been an active area of research and development over a twenty year period (Hoffer et al., 1975; Strahler, Logan, & Bryant, 1978). A new ARTMAP-based methodology for automatic mapping from Landsat Thematic Mapper (TM) and terrain data has been tested on a challenging remote sensing classification problem, using spectral and terrain features for vegetation classification in the Cleveland National Forest (Carpenter et al., 1997a). After training at the pixel level, system capabilities are tested at the stand level, using sites not seen during training. ARTMAP performance was compared to those of maximum likelihood classifiers, as well as back propagation neural networks and K Nearest Neighbor (KNN) algorithms. ARTMAP learning, being fast, stable, and scalable, overcomes common limitations of back propagation, which did not give satisfactory performance on this problem. Best results were obtained using a hybrid system based on a convex combination of fuzzy ARTMAP and maximum likelihood predictions. The prototype remote sensing example below (Section 4.1) introduces each aspect of data processing and fuzzy ARTMAP classification (Section 4.2). The example shows how the network automatically constructs a minimal number of recognition categories to meet accuracy criteria (Section 4.3). A voting strategy (Section 4.4) improves prediction by training the system several times on different orderings of an input set. Voting assigns confidence estimates to competing predictions.

### 4.1 A Prototype Remote Sensing Problem

The prototype remote sensing task is learning to identify one of three CALVEG (Matyas & Parker, 1980) vegetation classes (mixed conifer, coast live oak, southern mixed chaparral) for sites at which two spectral values (Landsat TM1 and TM4) are known at each pixel. The prototype example is based on a data set collected at the Cleveland National Forest. Larger scale simulations on this data set predict 8 possible vegetation classes with inputs of up to 6 TM bands and 7 ancillary variables. In this more realistic setting, fuzzy ARTMAP performance compares favorably with that of maximum likelihood (Lillesand & Kiefer, 1994, pp. 594-596; Richards, 1993), K Nearest Neighbor (Duda & Hart, 1973), and back propagation (Rumelhart, Hinton, & Williams, 1986; Werbos, 1974). However, reducing the number of input dimensions to $M = 2$ (TM bands) and the number of output classes to $L = 3$ (vegetation classes) allows visual illustration of fuzzy ARTMAP dynamics, as follows.

The data set for the prototype remote sensing problem reports the vegetation class for each of 50 sites: 16 mixed conifer, 25 coast live oak, and 9 southern mixed chaparral (Table 1a). The sites

vary in size, averaging about 90 pixels each. Landsat spectral bands TM1 and TM4 constitute the data set input for each pixel, with values scaled to the interval $[0,1]$. Before training, 10 sites, representative of the vegetation class mix, are reserved as a test set. No pixels from these sites are used during training. The goal is to predict the correct vegetation class label for each of the 10 test set sites.

During training and testing, a given pixel corresponds to an $\mathrm{ART}_a$ input $\mathbf{a} \equiv (a_1, a_2)$, where $a_1$ is the value of TM1 and $a_2$ is the value of TM4 at that pixel. The corresponding $\mathrm{ART}_b$ input vector **b** represents the CALVEG vegetation class of the pixel's site:

$$\mathbf{b} = \begin{cases} (1,0,0) & \text{mixed conifer} \\ (0,1,0) & \text{coast live oak} \\ (0,0,1) & \text{southern mixed chaparral .} \end{cases}$$

During training, vector **b** informs the ARTMAP network of the vegetation class to which the pixel's site belongs. This supervised learning process allows adaptive weights to encode the correct association between **a** and **b**. Simulations below examine the effect of training set size on predictive accuracy (Table 1b). To generate a training set of a given size, pixels are selected at random from the entire training set, which represents approximately 3600 pixels in 40 sites.

**Table 1:** Prototype remote sensing simulations

**a. Data set**

| **Class label** | **# sites** | **# pixels** |
|---|---|---|
| mixed conifer | 16 | 1336 |
| coast live oak | 25 | 2752 |
| southern mixed chaparral | 9 | 348 |
| TOTAL | 50 | 4436 |

**b. Fuzzy ARTMAP Incremental Learning**

| **Training set (# pixels)** | **Categories (# $F_2^a$ nodes)** | **Test set pixels (% correct)** | **Test set sites (# correct)** |
|---|---|---|---|
| 100 | 8 | 85.9% | 8/10 |
| 500 | 21 | 83.2% | 9/10 |
| 2000 | 72 | 88.5% | 10/10 |
| 3328 | 126 | 89.3% | 10/10 |

**c. Voting**

| **Input ordering (Figure 4)** | **Categories (# $F_2^a$ nodes)** | **Test set pixels (% correct)** | **Test set sites (# correct)** |
|---|---|---|---|
| (a) | 126 | 89.3% | 10/10 |
| (b) | 131 | 86.8% | 9/10 |
| (c) | 139 | 86.8% | 9/10 |
| (d) | 153 | 89.4% | 9/10 |
| (e) | 133 | 84.8% | 8/10 |
| average | 136 | 87.4% | 9/10 |
| voting | --- | 91.0% | 10/10 |

Other simulations show how voting can improve predictive accuracy (Table 1c). During testing, each test set pixel predicts a class, given the spectral band input values $a_1$ and $a_2$ for that pixel. Performance accuracy is measured both in terms of the percent of pixels that are correct and in terms of the fraction of sites that are correctly identified by a vote among pixels in the site.

The prototype remote sensing problem requires a trained network to predict the vegetation class (mixed conifer, coast live oak, or southern mixed chaparral) of a test set site, given TM bands 1 and 4 measured at each pixel in the site. This section illustrates fuzzy ARTMAP dynamics by showing how the network learns to make correct vegetation class predictions on this problem. Figure 3 illustrates why the problem is difficult: of the 4436 pixels in the data set (Table 1a), many share spectral band values within and between the three vegetation classes, and the three classes are not linearly separable. In fact the problem proved to be too difficult for back propagation to make accurate predictions.

During the initial learning phase, pixels are selected one at a time, at random, from the 40 training set sites. Fuzzy ARTMAP is trained incrementally, with each TM band vector **a** presented just once. Following a search, if necessary, the network selects an $\mathrm{ART}_a$ category by activating an $F_2^a$ node $J$ for the input pixel, then learns to associate category $J$ with the $\mathrm{ART}_b$ vegetation class $K$ of the site in which the pixel is located. With fast learning, the class prediction $K$ of each $\mathrm{ART}_a$ category $J$ is permanent. If some input **a** with a different class prediction later selects this category, match tracking will raise $\mathrm{ART}_a$ vigilance $\rho$ just enough to trigger a search for a different $\mathrm{ART}_a$ category.

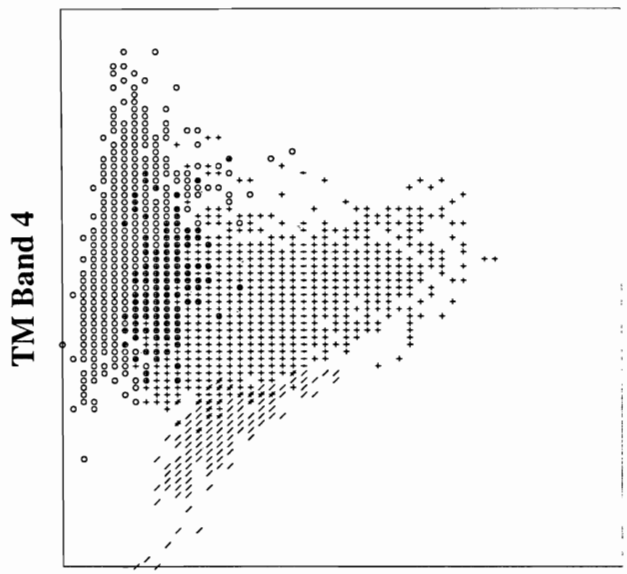


**Figure 3**: Prototype remote sensing inputs. Each point shows the scaled Landsat spectral band components $a_1$ (TM1 - blue) and $a_2$ (TM4 - near infrared) of the $\mathrm{ART}_a$ input vector **a**. Points o are found in mixed conifer sites, points + are found in coast live oak sites, and points / are found in southern mixed chaparral sites. Data set values are taken from the Cleveland National Forest.

### 4.2 Predictions of the Trained ARTMAP Network

As incremental learning proceeds, fuzzy ARTMAP creates a set of category nodes, each predicting one of the three vegetation classes. By the time 100 training set pixel inputs have been selected at random from the 40 training set site in a typical example, fuzzy ARTMAP has created 8 categories (Table 1b). Three of these categories predict mixed conifer, four predict coast live oak, and one predicts southern mixed chaparral. The 10 test set sites contain a total of 1108 pixels. After training on the first 100 inputs, network performance at this stage of learning was first measured by the number of correct vegetation class predictions the test set pixels were able to make. For each test set pixel, the TM band vector **a** selects one of the 8 $ART_a$ categories, then predicts that its site belongs to the vegetation class associated with that category. After training on just 100 input points, 85.9% of the test set pixels correctly predicted the vegetation classes of their sites. A second performance measure examined the number of test set sites that would be correctly classified. This method counts the number of pixels in each site that predict each vegetation class, then selects the class chosen by the most pixels. At this stage of learning, having used only 3% of the training set pixels, 8 of the 10 test site vegetation classes were correctly identified. In this case, too few southern mixed chaparral exemplars had been presented for that class to easily win a majority at any site.

As the number of training set inputs increased, the pixel-level predictive accuracy increased only marginally, even decreasing as the number of training set inputs increased from 100 to 500 (Table 1b). After presentation of all 3328 training set pixels, 89.3% of the test set pixels correctly predict the vegetation class of their site. However, site-level prediction improves steadily to 9/10 test set sites, after training on 500 inputs; and 10/10 sites, after training on 2000 inputs or on the full training set. This result highlights the observation that the pixel is often too small and noisy a unit to make an accurate prediction. However, a group of noisy pixel-level results can be pooled to form accurate mappings across functional regions or sites.

### 4.3 Voting

A typical characteristic of fast learning is dependence of category structure upon the order of training set input presentation. For example, suppose that two fuzzy ARTMAP networks learn from a common input set that is presented in two different orders during training. The two networks might then each correctly predict 90% of the test set inputs, despite the fact that the two have significantly different internal input grouping rules, or category boxes, at $ART_a$. In particular, the test set inputs that the first network identifies correctly are typically different from those that the second network identifies correctly, despite the fact that both were trained on the same input set. ARTMAP voting uses this order dependence to advantage to improve and stabilize overall predictive performance, as follows.

Figure 4a-e illustrates the decision regions of the prototype remote sensing example after presentation of all 3328 training set inputs (Table 1c). A decision region plot shows predictions all TM band inputs **a** would make if presented to the trained network. In Figure 3, data set points from mixed conifer sites were are represented by a circle (o), points from coast live oak sites by a plus (+), and points from southern mixed chaparral sites by a slash (/). The same marks indicate vegetation class predictions made by a network in response to spectral value inputs across the unit square. The rough decision boundaries in Figure 4a reflect the ambiguous predictions in the corresponding portion of the data set.

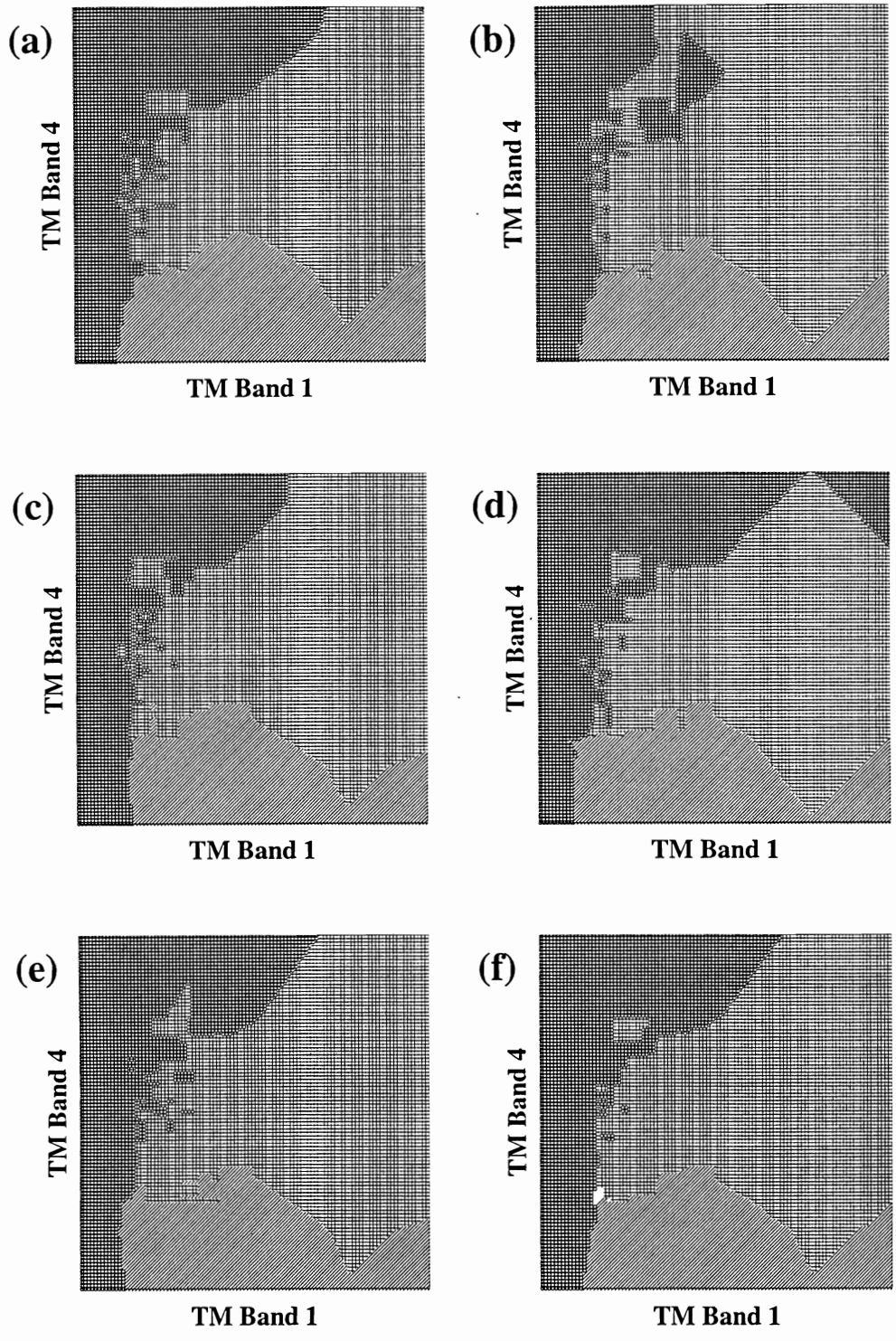


**Figure 4**: Prototype remote sensing example: Fuzzy ARTMAP voting. (a)-(e) Fuzzy ARTMAP networks trained on a common set of 3328 inputs presented in five different, random orders show variations in decision region geometry. Points marked by a circle (o) predict mixed conifer, points marked by a plus (+) predict coast live oak, and points marked by a slash (/) predict southern mixed chaparral. Pixel-level predictive accuracy ranges from 84.8% (e) to 89.4% (d) while site-level predictive accuracy ranges from 8/10 (e) to 10/10 (a) (Table 1c). (f) Voting across the five trained networks boosts pixel-level accuracy to 91.0% and site-level accuracy to 10/10. Blank spaces indicate a 2-2-1 tie among the voters.

Figure 4a-e and Table 1c show how network predictions can vary as a function of input order. Each of these five tests uses the same training set, presented in different, randomly chosen, orders. Decision boundaries vary, as do the number of $ART_a$ categories (from 126 to 153), the number of correct test set pixels (from 84.8% to 89.4%), and the number of correct test set site identifications (from 8/10 to 10/10). Before knowing the test set answers, it would be difficult to decide which of these five networks would be the most accurate on novel data. ARTMAP voting chooses for each pixel the class prediction chosen by the largest number of the five "voting committee" networks. The size of each vote also provides a measure of confidence in each decisions. Confidence is typically lowest near decision boundaries. Figure 4f indicates how voting can smooth and stabilize decision boundaries. In addition, pixel-level performance on the voting network (91.0%) is better than that of any individual trained network, and site-level prediction is perfect (10/10).

## 5 ARTMAP Variations for Applications

ART and ARTMAP networks feature winner-take-all (WTA) competitive coding, which groups inputs into disjoint recognition categories. Other neural network learning systems such as back propagation feature distributed coding, which can provide good noise tolerance and code compression but which typically requires slow learning. Fast learning tends to cause catastrophic forgetting in these networks, as it does in ART and ARTMAP networks in which the code representation is distributed. On the other hand, fast learning is often desirable for on-line adaptation to rapidly changing circumstances and for encoding of rare cases and large databases.

Variants of the basic ART and ARTMAP networks can acquire some of the advantages of distributed coding while maintaining fast learning capability. For example, ART-EMAP (Carpenter & Ross, 1993, 1995) uses WTA codes for learning and distributed codes for testing. Distributed prediction can significantly improve ARTMAP performance, especially when the size of the training set is small. In medical database prediction problems, which often feature inconsistent training input predictions, ARTMAP-IC (Carpenter & Markuzon, 1998) improves performance with a combination of distributed prediction, category instance counting, and a new match tracking search algorithm. A voting strategy further improves prediction by training the system several times on different orderings of an input set. Voting, instance counting, and distributed representations combine to form confidence estimates for competing predictions. However, since these and most other ART and ARTMAP variants use WTA coding during learning, they do not solve problems such as category proliferation with noisy training sets, unless learning is slow.

### 5.1 Distributed ART and Distributed ARTMAP

A new class of ART and ARTMAP models retain stable coding, recognition, and prediction, but allow arbitrarily distributed code representation during learning as well as performance (Carpenter, 1997; Carpenter, Milenova, & Noeske, 1998). These networks automatically apportion learned changes according to the degree of activation of each coding node. This permits fast as well as slow learning without catastrophic forgetting. Distributed ART models replace the traditional neural network path weight with a *dynamic weight* equal to the rectified difference between coding node activation and an adaptive threshold. The input signal $T_j$ that activates the distributed code is a function of a *phasic component* $S_j$, which depends on the active input, and a *tonic component* $\Theta_j$, which depends on prior learning but is independent of the current input. At each synapse, phasic and tonic terms balance one another and exhibit dual

computational properties. For example, during learning with a constant input, phasic terms are constant while tonic terms may grow. Tonic components would then become larger for all inputs, but phasic components would become more selective, reducing the total coding signal sent by a significantly different input pattern. Inputs activate distributed codes through phasic and tonic signal components with dual computational properties, and a parallel distributed match-reset-search process helps stabilize memory. When the code is winner-take-all, the unsupervised distributed ART model (dART) is computationally equivalent to fuzzy ART and the supervised distributed ARTMAP model (dARTMAP) is equivalent to fuzzy ARTMAP. With fast distributed learning, dART and dARTMAP networks are likely to further expand the domain of applications of the ART family of networks.

## REFERENCES

Bachelder, I.A., Waxman, A.M., & Seibert, M. (1993). A neural system for mobile robot visual place learning and recognition. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. **I**-512-517). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baloch, A.A., & Waxman, A.M. (1991). Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN. *Neural Networks*, **4**, 271-302.

Bernardon, A.M., & Carrick, J.E. (1995). A neural system for automatic target learning and recognition applied to bare and camouflaged SAR targets. *Neural Networks*, **8**, 1103-1108.

Carpenter, G.A. (1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks*, **10**, 1473-1494.

Carpenter, G.A., Gjaja, M.N., Gopal, S., & Woodcock, C.E. (1997a). ART neural networks for remote sensing: Vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geoscience and Remote Sensing,* **35**, 308-325.

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., & Woodcock, C.E., (1997b). A neural network method for mixture estimation for vegetation mapping. Submitted to *Remote Sensing of Environment.* Technical Report CAS/CNS TR-97-014, Boston, MA: Boston University.

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., & Franklin, J., (1998). A neural network method for efficient vegetation mapping. Submitted to *Remote Sensing of Environment.* Technical Report CAS/CNS TR-98-035, Boston, MA: Boston University.

Carpenter, G.A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54-115.

Carpenter, G.A., & Grossberg, S. (1987b). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26**, 4919-4930.

Carpenter, G.A. & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, **3**, 129-152.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698-713.

Carpenter, G.A., Grossberg, S., & Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565-588.

Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an Adaptive Resonance system. *Neural Networks*, **4**, 759-771.

Carpenter, G.A. & Markuzon, N. (1998). ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases. *Neural Networks*, **11**, 323-336.

Carpenter, G.A., Milenova, B., & Noeske, B. (1998). dARTMAP: A neural network for fast distributed supervised learning. *Neural Networks*, **11**, 793-813.

Carpenter, G.A., & Ross, W.D. (1993). ART-EMAP: A neural network architecture for learning and prediction by evidence accumulation. In *Proceedings of the World Congress on Neural Networks (WCNN'94)* (pp. **III** - 649-656). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carpenter, G.A., & Ross, W.D. (1995). ART-EMAP: A neural network architecture for object recognition by evidence accumulation. *IEEE Transactions on Neural Networks*, **6**, 805-818.

Caudell, T.P., & Healy, M.J. (1994). Adaptive Resonance Theory networks in the Encephalon autonomous vision system. In *Proceedings of the 1994 IEEE International Conference on Neural Networks* (pp. **II**-1235-1240). Piscataway, NJ: IEEE.

Caudell, T.P., Smith, S.D.G., Escobedo, R., & Anderson, M. (1994). NIRS: Large scale ART 1 neural architectures for engineering design retrieval. *Neural Networks*, **7**, 1339-1350.

Christodoulou, C.G., Huang, J., Georgiopoulos, M., & Liou, J.J. (1995). Design of gratings and frequency selective surfaces using fuzzy ARTMAP neural networks. *Journal of Electromagnetic Waves and Applications*, **9**, 17-36.

Dubrawski, A., & Crowley, J.L. (1994a). Learning locomotion reflexes: A self-supervised neural system for a mobile robot. *Robotics and Autonomous Systems*, 12, 133-142.

Dubrawski, A., & Crowley, J.L. (1994b). Self-supervised neural system for reactive navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, San Diego, May, 1994 (pp. 2076 - 2081). Los Alamitos, CA: IEEE Computer Society Press.

Duda, R.O., & Hart, P.E. (1973). Pattern Classification and Scene Analysis. New York: Wiley.

Gan, K.W., & Lua, K.T. (1992). Chinese character classification using an Adaptive Resonance network. *Pattern Recognition*, **25**, 877-88.

Gjerdingen, R.O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception*, **7**, 339-370.

Gopal, S., Sklarew, D.M., & Lambin, E. (1994). Fuzzy-neural networks in multi-temporal classification of landcover change in the Sahel. In *Proceedings of the DOSES Workshop on New Tools for Spatial Analysis*. Lisbon, Portugal, DOSES, EUROSTAT. ECSC-EC-EAEC: Brussels, Luxembourg, pp. 55-68.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187-202.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87**, 1-51.

Ham, F.M., & Han, S.W. (1993). Quantitative study of the QRS complex using fuzzy ARTMAP and the MIT/BIH arrhythmia database. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. **I**-207-211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hoffer, R.M., & Staff (1975). Natural resources mapping in mountainous terrain by computer analysis of ERTS-1 satellite data. Agricultural Experiment Station Research Bulletin 919, and LARS Contract Report 061575, W. Lafayette, IN: Purdue University, 124 pp.

Johnson, C. (1993). Agent learns user's behavior. *Electrical Engineering Times*, June 28, pp. 43, 46.

Kasperkiewicz, J., Racz, J., & Dubrawski, A. (1995). HPC strength prediction using artificial neural network. *Journal of Computing in Civil Engineering*, **9**, 279-284.

Koch, M.W., Moya, M.M., Hostetler, L.D., & Fogler, R.J. (1995). Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition. *Neural Networks*, **8**, 1081-1102.

Lillesand, T.M., & Kiefer, R.W. (1994). Remote Sensing and Image Interpretation. Third edition. New York: John Wiley.

Ly, S., & Choi, J.J. (1994). Drill condition monitoring using ART-1. In *Proceedings of the 1994 IEEE International Conference on Neural Networks* (pp. **II**-1226-1229). Piscataway, NJ: IEEE.

Matyas, W.J., and Parker, I. (1980). CALVEG mosaic of existing vegetation of California. San Francisco: Regional Ecology Group, US Forest Service, Region 5, 630 Sansome Street. 27 pp.

Mehta, B.V., Vij, L., & Rabelo, L.C. (1993). Prediction of secondary structures of proteins using fuzzy ARTMAP. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. **I**-228-232). Hillsdale, NJ: Lawrence Erlbaum Associates.

Murshed, N.A., Bortozzi, F., & Sabourin, R. (1995). Off-line signature verification, without *a priori* knowledge of class $\omega_2$. A new approach. In *Proceedings of ICDAR 95: The Third International Conference on Document Analysis and Recognition.*

Richards, J. (1993). Remote Sensing Digital Image Analysis: An Introduction (pp. 182-189). Springer-Verlag: Berlin.

Rumelhart, D.E., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), Parallel Distributed Processing (pp. 318-362). Cambridge, MA: MIT Press.

Seibert, M., & Waxman, A.M. (1992). Adaptive 3D object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 107-124.

Seibert, M., & Waxman, A.M. (1993). An approach to face recognition using saliency maps and caricatures. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. **III**-661-664). Hillsdale, NJ: Lawrence Erlbaum Associates.

Strahler, A.H., Logan, T.L., & Bryant, N.A. (1978). Improving forest cover classification accuracy from Landsat by incorporating topographic information. *Proceedings of the 12th International Symposium on Remote Sensing of Environment* (pp. 927-942). Ann Arbor, MI: Environmental Research Institute of Michigan.

Suzuki, Y., Abe, Y., & Ono, K. (1993). Self-organizing QRS wave recognition system in ECG using ART 2. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. **IV**-39-42). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tarng, Y.S., Li, T.C., & Chen, M.C. (1994) Tool failure monitoring for drilling processes. In *Proceedings of the 3rd International Conference on Fuzzy Logic, Neural Nets and Soft Computing* (pp. 109-111), Iizuka, Japan.

Waxman, A.M., Seibert, M.C., Gove, A., Fay, D.A., Bernardon, A.M., Lazott, C., Steele, W.R., & Cunningham, R.K. (1995). Neural processing of targets in visible, multispectral IR and SAR imagery. *Neural Networks*, **8**, 1029-1051.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. PhD Thesis, Cambridge, MA: Harvard University.

Wienke, D., & Kateman, G. (1994). Adaptive Resonance Theory based artificial neural networks for treatment of open-category problems in chemical pattern recognition - Application to UV-Vis and IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems.*

Wienke, D., Xie, Y., & Hopke, P.K. (1994). An Adaptive Resonance Theory based artificial neural network (ART 2-A) for rapid identification of airborne particle shapes from their scanning electron microscopy images. *Chemometrics and Intelligent Laboratory Systems.*