The Bayesian Hierarchical Classifier (BHC) and Its Application to Short Vegetation Using Multifrequency Polarimetric SAR

Yanni Kouskoulas, Fawwaz T. Ulaby, *Fellow, IEEE*, and Leland E. Pierce, *Senior Member, IEEE*

***Abstract*—Given an image of a scene comprised of a number of distinct terrain classes, the optimum Bayesian classifier (OBC) provides the highest possible classification accuracy of the imaged scene, provided we have *a priori* knowledge of the probability density function (pdf) of the sensor's output for each terrain class. If the imaging sensor consists of multiple channels, application of OBC requires knowledge of the joint pdf of the observations made by all the channels. In practice, the volume of data needed in order to generate an accurate multidimensional pdf far exceeds the size of available datasets. The data-size requirement may be relaxed by assuming the pdfs to be Gaussian in form, but such an assumption leads to suboptimum classification performance. This paper addresses the data size issue by 1) taking advantage of the maximum-entropy density estimation (MEDE) technique introduced in a companion paper and 2) using marginal pdfs in a hierarchical approach. Using multidate synthetic aperture radar observations, it was shown that the Bayesian hierarchical classifier introduced in this paper can classify short vegetation classes with an accuracy of 93%, without retraining, compared with an accuracy of 84% for the maximum-likelihood estimator (with Gaussian assumption) and only 74% with ISODATA.**

***Index Terms*—Adaptive estiamtion, image classification, maximum-entropy methods, probability.**

I. Introduction

In the context of remote sensing, level-1 terrain classification involves the use of remotely sensed images to establish the general-class identity of every pixel of the imaged scene. When the imaging sensor is a synthetic aperture radar (SAR), level-1 classes typically include urban, water, bare soil, tall vegetation (trees) and short vegetation (usually less than 3 m in height). The classification process involves the use of an algorithm, and at level 1, the classification accuracies of available algorithms are quite high for both multichannel optical imagery and for multipolarized SAR observations. The much more daunting challenge is that of level-2 classification, particularly of short vegetation. In that case, the goal is to separate different vegetation types from each other, which is inherently difficult because of the fact that short vegetation is a dynamic medium whose scattering properties vary widely over the growing season. To illustrate what we mean, let us examine data displayed in Fig. 1 which shows the range of values measured by an airborne radar for σ^0_{CHV} for one field each of corn, wheat, soybeans, and alfalfa. The quantity σ^0_{CHV} corresponds to the signal intensity that is reflected per unit area of ground for a single polarimetric channel (HV) at C-band, and is chosen here because it is useful for classification purposes and represents the characteristics of the typical radar measurement. A more in-depth discussion of the different quantities measured by the radar which are used for classification follows in Section II-A. These four fields are part of a test site that was imaged on six different occasions between May 31 and July 14, in 1995. The range of values for each date and crop type, portrayed in the form of a vertical line in Fig. 1, is associated primarily with spatial variations of the biophysical properties over nine spatially different segments of that field, each approximately 30 m $\times$ 30 m in size. Each data point contained in that interval is the average of 75 independent samples (produced by applying multilook SAR processing and subsequently spatially averaging homogeneous groups of 25 pixels together). This averaging was performed to reduce speckle related variations down to an insignificant level. As we can see from the data in Fig. 1, several of the vertical bars for a given date overlap each other, making discrimination error prone, and if we were to include vertical bars for data measured in all fields for a particular crop type (instead of for only one field), then the vertical bars would overlap further and degrade classification accuracy. Using multiple SAR

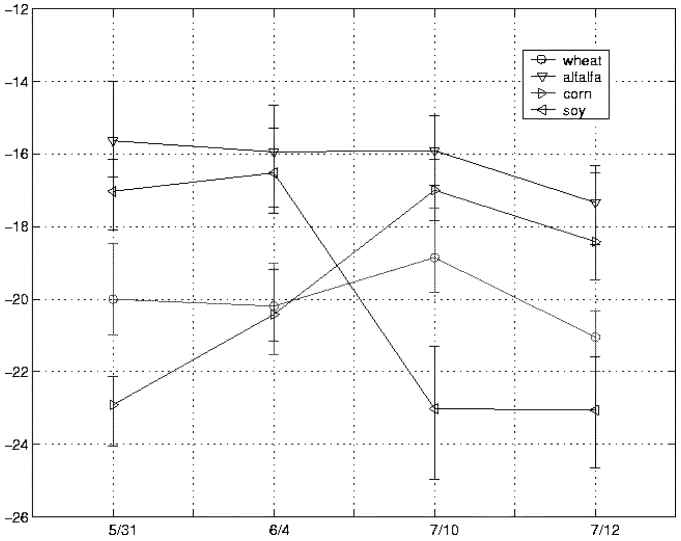


Fig. 1. Wheat, alfalfa, corn, and soybean radar measurements over four different days. Vertical bars indicate the dynamic range of measurements within a single field. This plot does not include the noise due to fading, which would increase the vertical bars significantly.

Manuscript received October 25, 2001; revised March 21, 2003.

The authors are with the the Radiation Laboratory, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: ulaby@umich.edu).

Digital Object Identifier 10.1109/TGRS.2003.821066

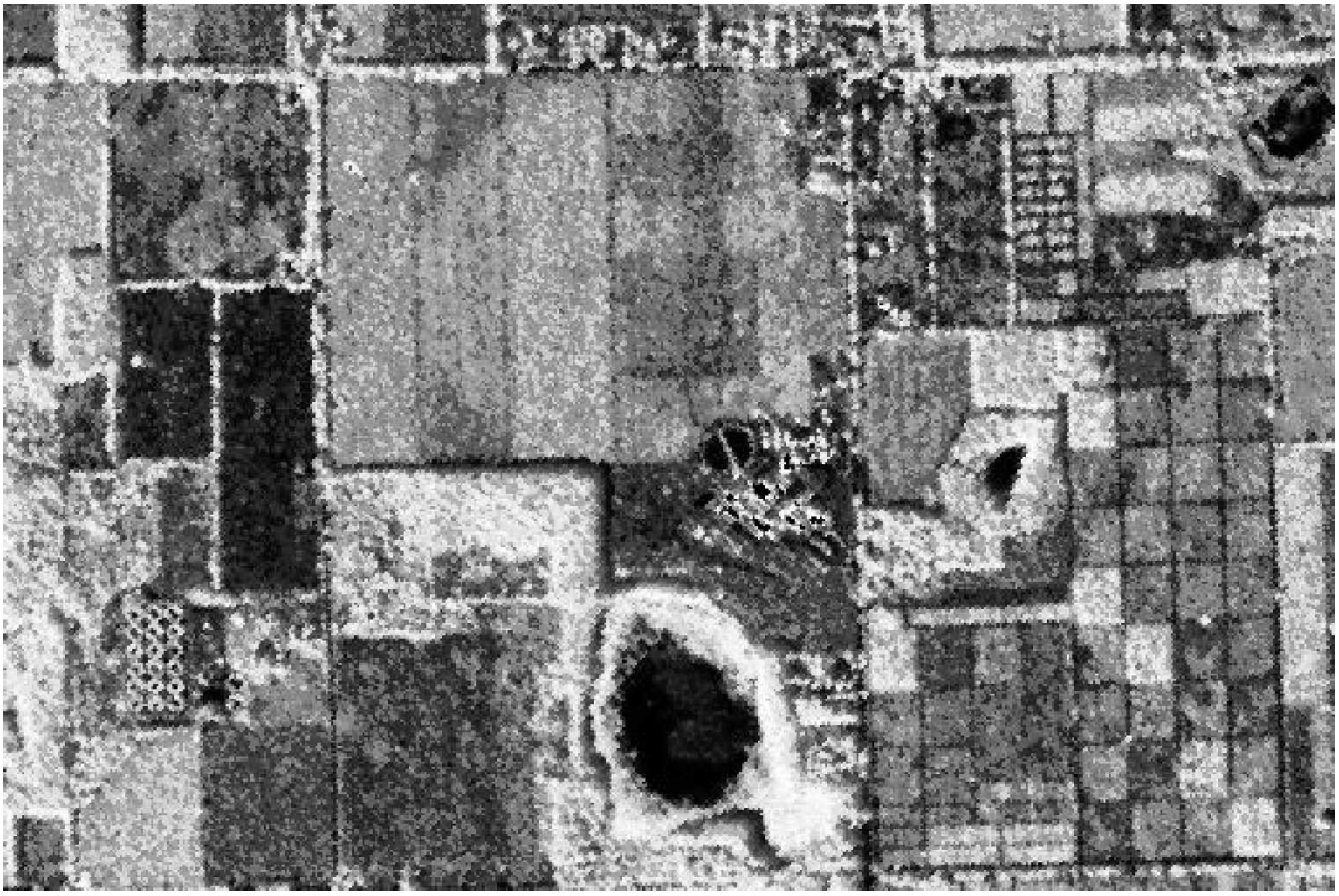


Fig. 2. SAR image of the test site at the Kellogg Biological Station. The square plots of homogeneous vegetation in the lower right of the image are 100 m on a side.

channels, there is a reasonable chance of developing a classifier through supervised training that will yield good classification accuracies when applied to independent fields observed in the same image on the same day. It is far more challenging to be able to develop an algorithm that can be applied to imagery acquired on any day during the growing season and still produce good accuracies, without retraining. That challenge is the very goal of the present study. That is, we seek to develop an algorithm that, subsequent to training, can be applied "blindly" to SAR imagery acquired in May, June, or July, or in the general case, at any time.

II. Experimental Dataset

Before we proceed with our stated task, we should provide more information on the sensor we have in mind and the dataset we plan to use. The sensor is a combined L- and C-band polarimetric radar imager similar to the SIR-C system that flew aboard the space shuttle. The data we plan to use was acquired by an airborne version of SIR-C, called AirSAR [1], mounted on a Jet Propulsion Laboratory aircraft. The AirSAR was used in 1995 to acquire SAR imagery of a site at the Kellogg Biological Station, located near Kalamazoo, MI, on six different dates extending over a six-week long period during the growing season. A sample image for one of the six dates is shown in Fig. 2. This image shows several rectangular plots in the lower right and the upper middle that will be used in this study.

A. SAR Measurement Channels

For each frequency band (L-band at $\lambda = 23$ cm and C-band at $\lambda = 6$ cm), the polarimetric SAR processor produces a two-dimensional (2-D) calibrated image in which each 5 m $\times$ 5 m pixel is characterized by a complex scattering matrix S, defined by

$$E^{\text{backscattered}} = \frac{1}{4\pi r^2} \begin{pmatrix} S_{VV} & S_{VH} \\ S_{HV} & S_{HH} \end{pmatrix} E^{\text{incident}} \quad (1)$$

where the four scattering matrix elements refer to the four combinations of linear polarization for the transmit and receive operations: $_{VH}$ refers to vertical polarization received and horizontal polarization transmitted. Each element is complex, in order to represent both amplitude and phase of the backscattered electric field.

To effectively remove speckle noise [12], [13], [23], [24], the boundaries of each agricultural field in the test site were delineated and then the field was partitioned into sections such that each section comprised on the order of 25 pixels. Upon averaging the 25 pixels in a section, thereby increasing the total number of independent samples to 75, signal fluctuations that

TABLE I
DATA SUMMARY

Dates in 1995: May 31, June 2, June 4, July 10, July 12, July 14
Sensor: JPL AirSAR
Site: Kellog Biological Station, Kalamazoo, MI, U.S.A.
Channels: L and C band, polarimetric data
Independent samples per data point: 70

Crop	# Points	Vegetation Heights	Vegetation Biomass	Volumetric Soil Moisture
Alfalfa	1845	10-60cm	0.01-5.6 kg/m^2	0.14-0.33 cm^3/cm^3
Corn	864	93-118cm	1.3-2.0 kg/m^2	0.11-0.17 cm^3/cm^3
Wheat	1260	49-103cm	1.6-3.1 kg/m^2	0.14-0.35 cm^3/cm^3
Bare	909	0cm	0 kg/m^2	0.15-0.18 cm^3/cm^3
Soy	360	$\star$	$\star$	$\star$

manifest themselves in the form of a speckled appearance in the image were essentially filtered out. The averaging process was performed on the elements of the covariance matrix of S which were then used to generate the following five quantities:

$$\sigma^o_{HH} = \frac{4\pi}{A}\left\langle |S_{HH}|^2 \right\rangle \tag{2}$$

$$\sigma^o_{VV} = \frac{4\pi}{A}\left\langle |S_{VV}|^2 \right\rangle \tag{3}$$

$$\sigma^o_{HV} = \frac{4\pi}{A}\left\langle |S_{HV}|^2 \right\rangle \tag{4}$$

$$\alpha = \frac{|\langle S_{VV} S^*_{HH} \rangle|}{\sqrt{\langle |S_{VV}|^2 \rangle \langle |S_{HH}|^2 \rangle}} \tag{5}$$

$$\zeta = \tan^{-1} \frac{\langle \Im(S_{VV} S^*_{HH}) \rangle}{\langle \Re(S_{VV} S^*_{HH}) \rangle} \tag{6}$$

where A is the pixel area. The first three quantities are the linearly polarized scattering coefficients, which measure the overall strength of the electromagnetic reflection for that polarization combination and for that pixel, while the last two describe the statistics of the phase of the reflected wave for that pixel, giving the mean and the spread of the phase statistics. The quantity α is known as the copolarized phase correlation coefficient, while ζ is the mean value of the copolarized phase difference [19]. Examination of these five quantities revealed that ζ offered little, if any, discriminating power in the crop-classification problem. The measured scattering matrices can not only be used as above to generate estimates of α and ζ using the combination of a copolarized scattering amplitude, such as S_{VV}, but also can be used to compute the cross polarized scattering amplitude using S_{VH}. These cross-polarized phase attributes were also found to offer no discrimination power for vegetation classes. Hence, only the three scattering coefficients and the phase correlation coefficient α were used in the algorithm development. Thus, our measurement data vector consists of four dimensions per band, or a total of eight dimensions altogether.

B. Test Site Data

Table I provides a summary of the *in situ* measurements made in the field in support of the six overflights. For each cover type the entry labeled "# points" refers to the number of 5 m $\times$ 5 m sections for which radar data was extracted over the six flights and for which (in most cases) *in situ* measurements were recorded. The following are the key points we should underscore.

1) Our available database of over 5000 ground sections is very large by the standard of any crop classification study conducted to date.
2) Those ground sections consist of five classes (four crop covers and bare soil) each of which spans a wide range of biophysical parameters and underlying soil-surface properties.

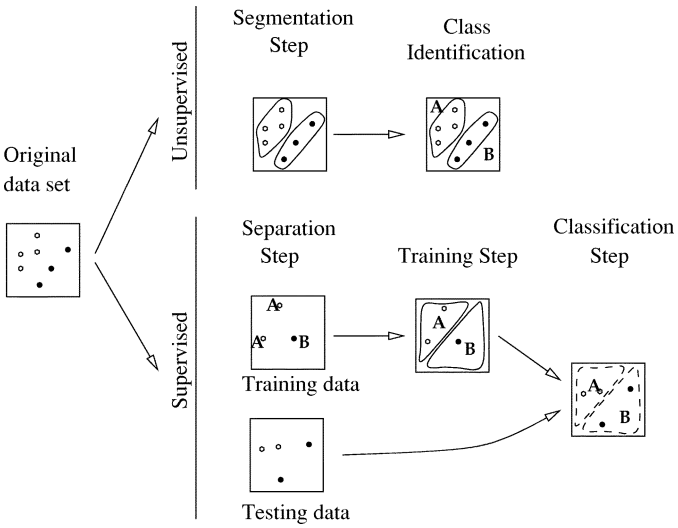


Fig. 3. Difference between unsupervised and supervised classifiers.

III. OVERVIEW OF CLASSIFIER TECHNIQUES

The literature for radar-based classifiers usually categorizes classification algorithms into two different types: supervised and unsupervised [6]. An *unsupervised* classifier is given raw data, and the classification algorithm separates data into clusters, without any extra information. Each cluster must subsequently be identified as a class by hand. A *supervised* classifier requires more preparation; it must be given a set of data with corresponding classes (usually called training data) with which it tries to "learn" the character of that particular class. After this step, it uses what it learned and applies it to classifying any other data that it is given. This is diagrammed in Fig. 3

There are a wide variety of approaches to the classification problem, and a varied set of tools that have been developed to aid in different parts of the solution. Examples of some common tools and techniques that have been used for classification are: the method of principal components, Markov random field models, maximum-likelihood estimation (MLE), maximum *a posteriori* estimation (MAP), optimal Bayesian classification, minimum-distance classification, and parallelepiped classification. Classification algorithms are frequently built on combinations of these (or similar) tools.

Existing literature includes techniques based on different characteristics of covariance matrix elements [3], [11], [18], Markov random field models [21], unsupervised or partially supervised clustering techniques [2], [4], [7], neural networks [5], measures of texture [14], and polarimetric filtering [15], [20], [22], [25].

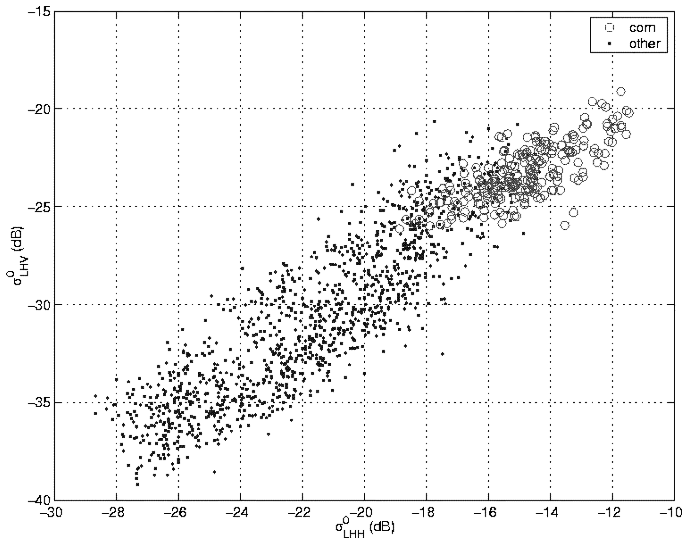


Fig. 4. First step of a multidimensional hierarchical decision rule is to select the pair of dimensions that provides the maximum separability, between the specific crop class and all the other classes.

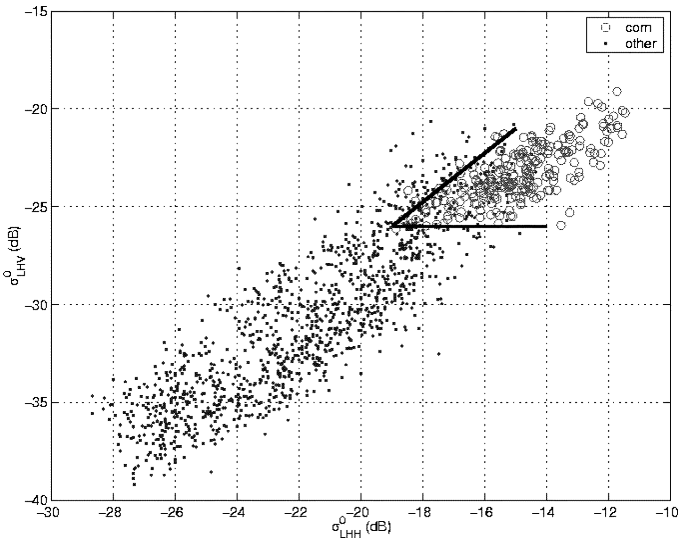


Fig. 5. Second step is to draw boundaries separating points that are exclusively noncorn from those that may be corn or not.

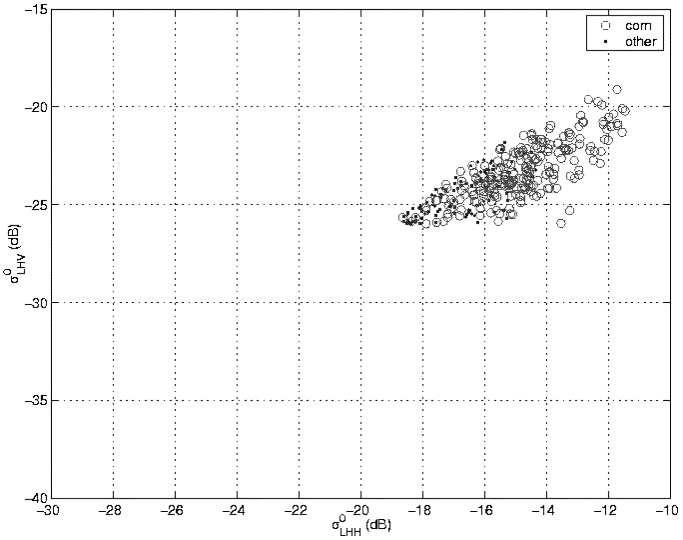


Fig. 6. Third step is to discard data that are in the region of exclusively noncorn. Note that the remaining data still include data from other classes mixed in with the corn data.

Some successful ideas, such as neural networks and optimal polarization, are excellent for certain types of classification problems, but do not work well on the problem at hand. Both techniques have difficulties because they do not adequately take into account the variation in measurements that we see over time for a single vegetation type. Other techniques, such as using texture measures and Markov random field models, are useful because they use the spatial information available in an image to help solve the classification problem. However, our investigation focuses only on the electromagnetic scattering characteristics of each pixel in isolation, so we will not try to exploit information about the spatial correlation information of the data.

Supervised methods in the literature are particularly useful because once trained, they are capable of producing completely automated classifications. Unsupervised methods require human intervention after the clustering phase, and, thus, are less desirable. In addition, methods such as ISODATA [2] have poor accuracy when applied to classification of short vegetation because they incorrectly group the clusters together, erroneously lumping one class within another.

Our goal is to develop an automatic supervised classifier capable of operating with multidimensional data and then to use it to generate accurate maps of vegetation classes at any time in the growing season, without retraining.

IV. Review of the Hierarchical Classifier

As described in [8], [10], and [16], the hierarchical classifier uses sequential decision rules based on training data, with each decision rule involving two of the eight available channels at a time. Instead of dealing with all of the classes simultaneously, the classifier deals with each vegetation class in a separate process. The process begins by examining all possible combinations of vegetation classes and channel-pairs in the form of 2-D scatter plots similar to that shown in Fig. 4, with the data in each belonging to only two categories, namely one of the vegetation classes as one of the categories and all the other classes as the second. From among these possibilities, we choose the vegetation type and specific channel combination that appears to separate that class from all others with the least confusion (overlap). In the present example, the combination of HV and HH polarizations, both at L-band, provided better separation for corn from the other six categories (short alfalfa, full alfalfa, bare soil, soybeans, early wheat, and late wheat) than any other combination of channels did for corn or for any of the other vegetation classes.

Next, we draw a boundary, such that on one side of the boundary all of the data are definitely not corn. The data points that remain on the other side of the boundary are a mixture of both corn, as well as data from the other vegetation classes. Fig. 5 shows the boundary we have drawn, and in Fig. 6, the data points to the left of the boundary have been discarded, which we know are not corn. The data that remain in Fig. 6 are then reprojected onto two different dimensions, and the

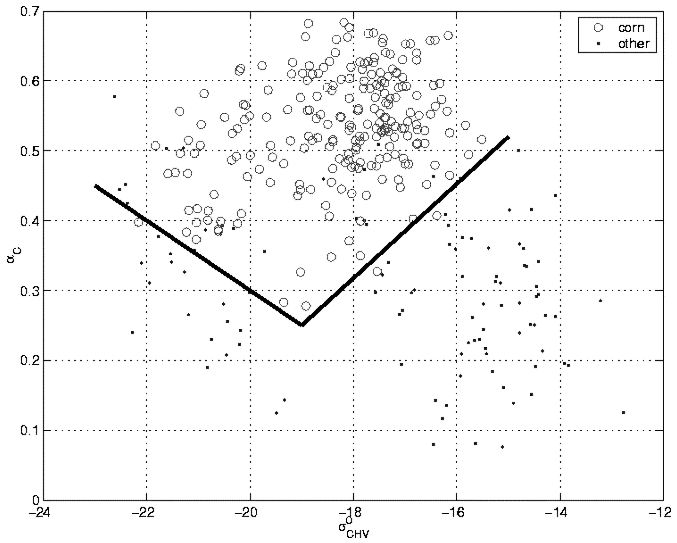


Fig. 7. Fourth step is to reproject the remaining data onto a new set of dimensions. This exposes noncorn regions and allows us to draw a new set of boundaries.

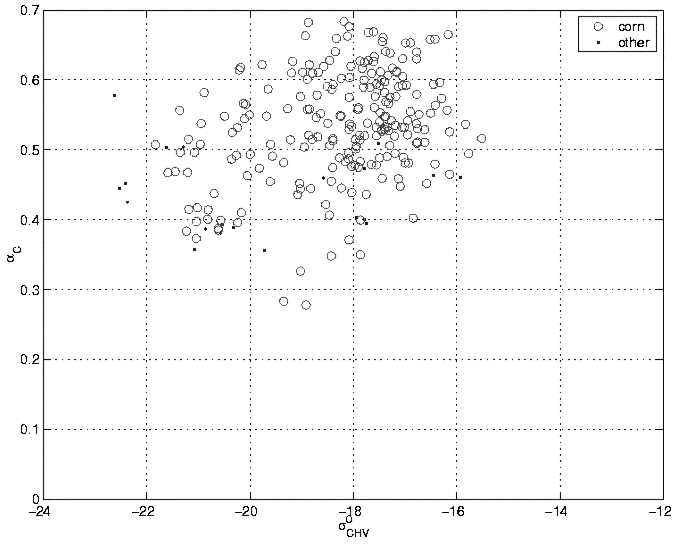


Fig. 8. Fifth step is to discard measurements in the exclsively noncorn region of the plot. This process can be repeated with as many projections as are useful.

process is repeated. Reprojecting our data into different pairs of dimensions is akin to looking at the multidimensional clouds of data from different angles. Each additional view we get can give us new information. Fig. 7 shows the results of reprojecting the radar measurements from Fig. 6 onto two new dimensions, namely α at C-band and the cross-polarized σ^o at C-band, and drawing a second set of boundaries. In Fig. 8, the noncorn side of the boundary is discarded. What remains after we are done is the set of points that are always in the region of our higher dimensional measurement space that contains our class of interest (in this case, corn). We have, piece by piece, defined a multidimensional boundary.

In the preceding paragraphs, we have described the process of drawing boundaries and how this allows us to separate corn from other classes. This process is repeated on the noncorn data that remains, for each additional vegetation class we wish to identify, and the resulting boundaries are applied to independent dataset.

TABLE II
ACCURACY OF HIERARCHICAL CLASSIFICATION USING INDEPENDENT TESTING DATA

	Classified As					
True Class	Alfalfa	Bare Soil	Corn	Soybeans	Wheat$^{(E)}$	Wheat$^{(L)}$
Alfalfa$^{(S)}$	88.4	1.7	2.2	3.2	0.2	4.3
Alfalfa$^{(T)}$	81.4	8.3	0.0	0.3	4.2	5.8
Bare	6.2	80.9	4.5	0.0	6.9	1.5
Corn	0.0	3.8	94.8	0.0	1.3	0.2
Soy	12.0	0.0	0.0	86.5	0.0	1.5
Wheat$^{(E)}$	0.7	3.0	3.0	0.0	86.3	7.0
Wheat$^{(L)}$	14.5	2.5	0.4	1.3	2.0	79.3

The results are presented in Table II as a confusion matrix. Each entry in the confusion matrix is a percentage value, indicating what percent of the data in the class labeled on the left side of the row was classified with the label in the column above. This is the convention we use for all of the confusion matrices in this text.

The accuracy measure that we present with the confusion matrices is known in the literature as user accuracy, which for a particular class C_i measures what percentage of the data identified as class C_i indeed belong to class C_i. This measure should not be confused with producer accuracy, which for class C_i measures what percentage of the data from class C_i was correctly classified, or overall accuracy, which is the percentage of all data that were correctly classified.

V. BHC

The natural extension of the hierarchical classifier would be to find an automated method of smoothly defining the multidimensional boundaries that contain each dataset. Because the regions in which the data are contained are probabilistic, we could improve our representation further if, instead of multidimensional shells, we used a multidimensional probability density function (pdf) to represent the region of data in each class. Once one has the pdfs representing the statistical variation of the vegetation class in measurement space, optimal Bayesian classification is the provably optimal method of classification [17]. OBC is usually formulated as

$$P(C_i|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|C_i)P(C_i)}{\sum_j f(\boldsymbol{x}|C_j)P(C_j)} \tag{7}$$

where $P(C_i|\boldsymbol{x})$ is the probability that we are measuring the ith class, C_i if the value of our measurement is $\boldsymbol{x}$. We can compute this probability if we know $f(\boldsymbol{x}|C_i)$, the probability that we measure $\boldsymbol{x}$ if we know we are looking at class C_i. We may also assume that we have no additional knowledge about the frequency of the classes (with N classes, that would mean that $P(C_i) = (1/N)$).

The reason no one has applied OBC to this problem is that it is difficult to accurately estimate a probability density function for an arbitrary dataset in multiple dimensions. We will address this difficulty in three ways: 1) we will use the the maximum-entropy density estimation (MEDE) technique, which works well with very sparse datasets; 2) we will construct this conditional

density with marginals, assuming that they are uncorrelated; and 3) we will use (7) in a hierarchical fashion similar to what was presented in Section IV, identifying one class at a time as opposed to all the classes at once. Using marginals allows us to perform density estimation on a subset of the eight dimensions, while still preserving important higher dimensional correlations needed to classify the data with high accuracy. We call this approach the Bayesian hierarchical classification (BHC) technique.

A. MEDE

MEDE (described in [9]) is based on the fact that using information theory, one can derive a parsimonious functional form for expressing density functions.

B. Classifier Implementation

The first step is training the classifier, and we begin by finding, for each class of interest (denoted C_i), marginal pdfs with which to approximate the overall conditional pdf. When combined into a single conditional pdf by taking their product, these marginals should be such that they best separate the class of interest from all the others. In our case, we can express a single data point as $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)$. If C_j denotes the jth class, then we would like to express the conditional pdf that the sensor measures $\boldsymbol{x}$, given that the sensor is observing a member of class C_j as

$$f(\boldsymbol{x}|C_j) = f(x_a, x_b, x_c|C_j) \cdot f(x_d, x_e, x_f|C_j) \cdot f(x_g, x_h|C_j) \tag{8}$$

where the set $(x_a, x_b, \ldots, x_h)$ maps on-to-one onto the set $(x_1, x_2, \ldots, x_8)$, with the exact mapping being determined by the procedure below. The dimensionality of the marginal pdfs must be chosen so that the available density of data is sufficient to perform the density estimates using the MEDE technique. In the case of this paper, the data volume listed in Table I is sufficient to generate three-dimensional (3-D) pdfs for all of the crop types listed in the table, and only marginally so for soybeans.

Approximating the density $f(\boldsymbol{x}|C_j)$ as the product of three marginal pdfs requires statistical independence between the groups of channels comprising the three marginal pdfs. The reality is that none of these channels, or groups of channels, is independent of the others. We can choose which channels to group together in each of the marginal pdfs such that the correlation between the groups is minimal, but that is not the approach we adopted in this paper. We decided instead to take a pragmatic approach as described in the procedure that follows.

We now turn our attention to the procedure for selecting the values of a, b, c, d, e, f, and g from among the set $(1, \ldots, 8)$, for each stage of the classification process. One can write out the steps as a cookbook procedure.

1) Use the training data and the MEDE to generate pdf estimates for each combination of three channels, for each class of data. For L channels and M classes, and d dimensions in each pdf, this corresponds to generating a total of $N = M\binom{L}{d}$ pdfs. In our case, $N = 280$ 3-D pdfs.
2) Perform classification using (7) N times, once for each combinations of three channels, and for each class, noting the resulting classification accuracy.
3) Choose the combination of three channels and the class (call it C_i) which yield the highest classification accuracy. This defines a, b, and c. Classification accuracy is defined as the average of user and producer accuracies for that category.
4) For C_i, choose the next most accurate combination of three channels that does not contain any of the first three channels. This defines d, e, and f.
5) Finally, the remaining two channels are grouped together, defining g and h.
6) Steps 1)–5) are then repeated sequentially for the remaining classes.

It is worth noting that for the training data provided to the classifier, wheat was divided into two subclasses: early wheat and late wheat. Similarly, alfalfa was divided into short and tall alfalfa. The accuracies in step two were computed on the basis of separate pdfs for each subclass, but after this step and during testing, the distinction of the subclasses was not used.

Once trained, the classifier is used in a hierarchical fashion, based on the approach described in Section IV. Starting from the first class of interest, (7) is used to with the appropriate set of marginals that best separates that class of interest from all the rest. Data points identified as the class of interest are so labeled. The next set of marginals is then applied to (7) to separate the second class of interest from the rest, using the remaining data. This process is repeated until all of the data have been classified.

TABLE III
CONFUSION MATRIX FOR SHORT VEGETATION PRODUCED WITH THE BHC. THE DATASET INCLUDED ALL SIX IMAGES THAT SPAN THE GROWING SEASON

	Classified As				
True Class	Alfalfa	Bare	Corn	Soy	Wheat
Alfalfa	90	3	0	2	5
Bare	4	90	3	0	3
Corn	0	0	99	0	1
Soy	4	0	0	93	3
Wheat	4	1	0	2	93

C. Results and Discussion

In this section, we will carefully evaluate the BHC using three different measures of accuracy, namely: user accuracy, producer accuracy, and overall accuracy. For each evaluation of accuracy for the BHC, we divided our dataset into two sets: a training dataset and a testing dataset. We used the training dataset to estimate the densities involved in our classification, and we used the testing dataset as an independent dataset with which to evaluate the performance of the classifier.

To begin our evaluation of classifier accuracy, we present in Table III a confusion matrix evaluating the performance of the BHC for the dataset we have available. Note that the number at the top of the confusion matrix is the producer accuracy averaged over all the classes, which gives an unbiased metric of classifier performance. For purposes of comparison, we classified the same data with ISODATA [2] and with an MLE classifier that assumes all data pdfs to be Gaussian in shape. The

TABLE IV
CONFUSION MATRIX BASED ON THE ISODATA UNSUPERVISED CLASSIFIER

	Classified As				
True Class	Alfalfa	Bare	Corn	Soy	Wheat
Alfalfa	62	8	0	2	28
Bare	10	73	7	0	10
Corn	1	16	81	0	3
Soy	4	1	0	74	22
Whcat	15	1	0	1	82

TABLE V
CONFUSION MATRIX BASED ON THE MLE CLASSIFIER WITH GAUSSIAN ASSUMPTION

	Classified As				
True Class	Alfalfa	Bare	Corn	Soy	Wheat
Alfalfa	76	2	2	5	16
Bare	9	68	17	0	7
Corn	0	6	92	0	2
Soy	3	0	0	96	0
Wheat	9	2	0	0	88

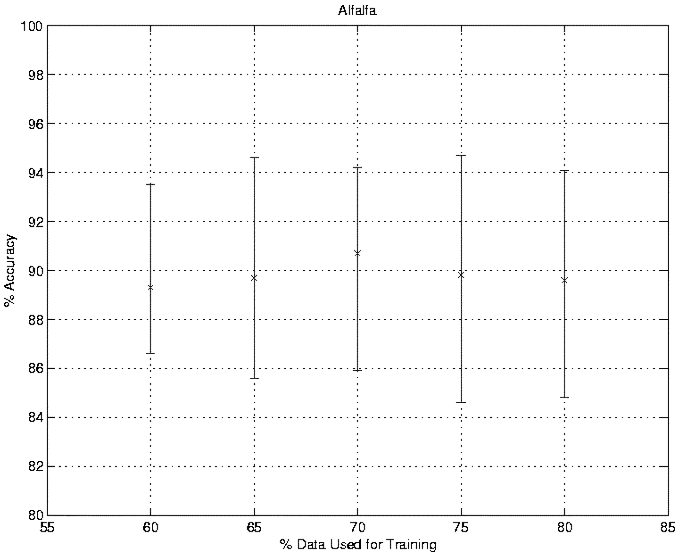


Fig. 9. User accuracy for alfalfa. The horizontal axis is the percent of data used for training, and the statistics were gathered over 50 independent trials. The vertical bars extend between the maximum and minimum accuracies, with the marks representing the mean accuracy.

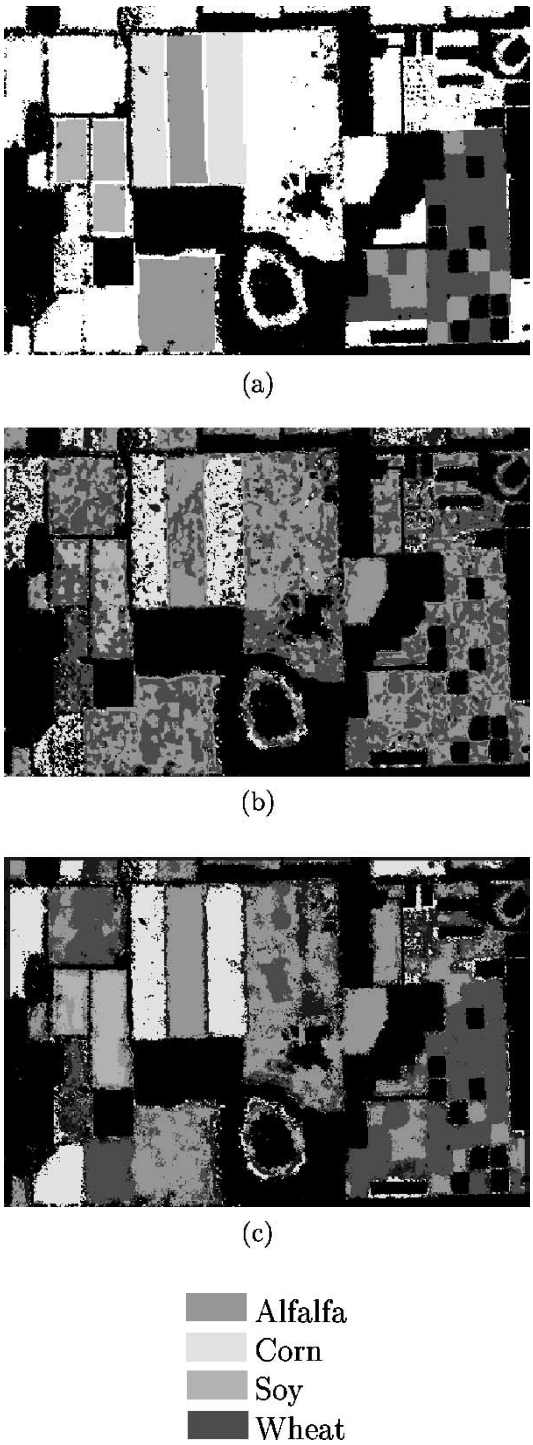


Fig. 10. (a) Ground cover information, (b) classification using ISODATA, and (c) classification using the BHC.

results are summarized in Tables IV and V, respectively. The comparison clearly demonstrates that the BHC is superior to the other classification methods, offering results with an average producer accuracy on the order of 10% higher than the results from the MLE classifier, and 20% higher than those associated with ISODATA. Note that while six images from throughout the growing season were used, the current method is not a multitemporal classifier: each image was classified independent of all the others and without knowledge of its acquisition date.

D. Classifier Accuracy Versus Size of Training Dataset

Because the results of the classifier depend heavily on the data used to train it, we randomized the selection of training and testing sets of data. The entire classification procedure, from training and density estimation to classification of the independent dataset, was performed 100 times. No data were used for both training and testing.

The results of these trials are summarized in Fig. 9 for alfalfa. In general, while the mean value of the user accuracy remained consistent, as the percentage of data used for training was increased between 60% and 80%, the range of values (minimum to maximum) of producer accuracy increased. This is attributed to the small size of the testing data. Similar trends were noted for the other vegetation classes.

E. Image Domain

Another evaluation of classifier accuracy was performed by examining results in the image domain. We applied the BHC to the radar image recorded on July 12, 1995, and we also applied ISODATA, an unsupervised clustering technique commonly found in the literature to classify the same image. The ground truth and the resulting classifications are compared in Fig. 10. Although neither classification is perfect, the reader will note the relative lack of accuracy in the ISODATA classification, which confuses wheat with alfalfa in many instances in the lower right corner of the image. In general, the homogeneous field sites are classified as random mixtures by ISOCLUS and as nearly homogeneous by our technique, as desired.

VI. Conclusion

Using eight channels of SAR data taken over a test site in Michigan, the Bayesian hierarchical classifier introduced in this paper was shown to correctly identify five vegetation cover classes with an accuracy of 93%. It is particularly noteworthy that once trained, the BHC was applied to classify images observed on six different dates covering a time span of six weeks, without retraining. Both the vegetation covers and the underlying soil moisture conditions varied widely over this time period.

The realized high classification accuracy is a tribute to both the new classification technique as well as to multifrequency multipolarization SAR as an effective mapper of terrain.

References

[1] AIRSAR Integrated Processor Documentation: Data Formats Version 0.01, Apr. 1995.

[2] G. H. Ball and D. J. Hall, "Isodata, A novel method of data analysis and pattern classification," Stanford Res. Inst., Stanford, CA, Tech. Rep., 1965.

[3] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 68–78, Jan. 1997.

[4] C. Fraley and A. E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis," Dept. Statistics, Univ. Washington, Seattle, WA, Tech. Rep. 329, Feb. 1998.

[5] Y. Ito and S. Omatu, "Polarimetric SAR data classification using competitive neural networks," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2665–2684, 1998.

[6] J. R. Jensen, *Introductory Digital Image Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[7] B. Jeon and D. A. Landgrebe, "Partially supervised classificatoin using weighted unsupervised clustering," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, Mar. 1999.

[8] Y. Kouskoulas, L. Pierce, F. Ulaby, and M. C. Dobson, "Classification of short vegetation using polarimetric SAR data," in *Proc. IGARSS*, vol. 1, 1999, pp. 103–105.

[9] Y. Kouskoulas, L. E. Pierce, and F. Ulaby, "A computationally efficient, multivariate maximum-entropy density estimation (MEDE) technique," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, pp. TBD–TBD, Feb. 2004.

[10] Y. Kouskoulas, F. Ulaby, and M. C. Dobson, "Classification of short vegetation using multifrequency SAR," *Proc. IEEE*, vol. 1, pp. 103–105, 1998.

[11] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.

[12] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R.Allen R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1017–1028, Sept. 1994.

[13] D.Donald Lewinski, "Nonstationary probabilistic target and clutter scattering models," *IEEE Trans. Antennas Propagat.*, vol. 31, pp. 490–498, May 1983.

[14] F. P. Miranda, L. E. N. Fonseca, J. R. Carr, and J. V. Taranik, "Analysis of JERS-1 (Fuyo-1) SAR data for vegetation discrimination in northwestern Brazil using the semivariogram textural classifier," *Int. J. Remote Sens.*, vol. 17, no. 17, pp. 3523–3529, 1996.

[15] T. Nagai, Y. Yamaguchi, and H. Yamada, "Use of multi-polarimetric enhanced images in SIR-C/X-SAR land-cover classification," *IEICE Trans. Commun.*, vol. E80-B, no. , Nov. 1997.

[16] L. E. Pierce, F. Ulaby, K. Sarabandi, and M. Craig Dobson, "Knowledge-based classification of polarimetric SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1081–1086, Sept. 1994.

[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 1992.

[18] E. Rignot, R. Chellapa, and P. Dubois, "Unsupervised segmentation of polarimetric SAR data using the covariance matrix," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 697–705, July 1992.

[19] K. Sarabandi, "Derivation of phase statistics from the Mueller matrix," *Radio Sci.*, vol. 27, no. 5, pp. 553–560, Sept.–Oct. 1992.

[20] K. Sarabandi and E. Li, "Characterization of optimum polarization for multiple target discrimination using genetic algorithms," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1810–1817, Dec. 1997.

[21] A. H. Schistad Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 100–113, Jan. 1996.

[22] A. A. Swartz, H. A. Yueh, J. A. Kong, L. M. Novak, and R. T. Shin, "Optimal polarizations for achieving maximum contrast in radar images," *J. Geophys. Res.*, vol. 93, no. B12, Dec. 1988.

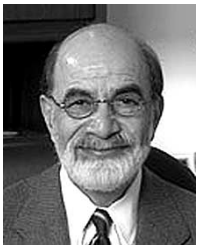
[23] F. T. Ulaby and C. Elachi, Eds., *Radar Polarimetry for Geoscience Applications*. Norwell, MA: Artech House, 1990.

[24] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing Active and Passive*. Norwell, MA: Artech House, 1982, vol. 2nd, pp. 476–483.

[25] J. van Zyl, "On The importance of polarization in radar scattering problems," Ph.D. thesis, California Instit. Technol., Pasadena, CA, 1986.

Yanni Kouskoulas received the B.S., M.S. and Ph.D. degrees from the University of Michigan, Ann Arbor, in 1995, 1996, and 2001, respectively.

He is currently with the Applied Physics Laboratory, Johns Hopkins University, Laurel, MD. His research interests include remote sensing, communications systems, and applied statistics.



Fawwaz T. Ulaby (M'68–SM'74–F'80) received the B.S. degree in physics from the American University of Beirut, Lebanon, in 1964, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Texas, Austin, in 1966 and 1968, respectively.

He is the Vice President for Research and Williams Distinguished Professor of electrical engineering and computer science at the University of Michigan, Ann Arbor. His current reserarch interests include microwave and millimeter-wave remote sensing, radar systems, and radio wave propagation. He has authored ten books and published more than 500 papers and reports on these subjects.

Dr. Ulaby is the recipient of numerous awards, including the Eta Kappa Nu Association C. Holmes MacDonald Award as "An Outstanding Electrical Engineering Professor in the United States of America for 1975," the IEEE Geoscience and Remote Sensing Distinguished Achievement Award in 1983, the IEEE Centennial Medal in 1984, The American Society of Photogrammetry's Presidential Citation for Meritorious Service in 1984, the NASA Group Achievement Award in 1990, and the 2000 IEEE Electromagnetics Award. He was President of the IEEE Geoscience and Remote Sensing Society from 1980 to 1982, Executive Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 1983 to 1985, and as General Chairman of several international symposia. In 1995, he was elected to membership in the National Academy of Engineering and currently serves as Editor-in-Chief of the PROCEEDINGS OF THE IEEE.

Leland E. Pierce (S'85–M'89–SM'01) received the B.S. degrees in both electrical and aerospace engineering in 1983, and the M.S. and Ph.D. degrees in electrical engineering in 1986 and 1991, respectively, all from the University of Michigan, Ann Arbor.

Since 1991, he has been the Head of the Microwave Image Processing Facility within the Radiation Laboratory, Electrical Engineering and Computer Science Department, The University of Michigan, where he is responsible for research into the uses of Polarimetric SAR systems for remote sensing applications, specifically forest canopy parameter inversion.