

## An AVHRR Multiple Cloud-Type Classification Package

PAUL M. TAG, RICHARD L. BANKERT, AND L. ROBIN BRODY

*Naval Research Laboratory, Monterey, California*

(Manuscript received 2 September 1998, in final form 22 March 1999)

### ABSTRACT

Using imagery from NOAA's Advanced Very High Resolution Radiometer (AVHRR) orbiting sensor, one of the authors (RLB) earlier developed a probabilistic neural network cloud classifier valid over the world's maritime regions. Since then, the authors have created a database of nearly 8000  $16 \times 16$  pixel cloud samples (from 13 Northern Hemispheric land regions) independently classified by three experts. From these samples, 1605 were of sufficient quality to represent 11 conventional cloud types (including clear). This database serves as the training and testing samples for developing a classifier valid over land. Approximately 200 features, calculated from a visible and an infrared channel, form the basis for the computer vision analysis. Using a 1-nearest neighbor classifier, meshed with a feature selection method using backward sequential selection, the authors select the fewest features that maximize classification accuracy. In a leave-one-out test, overall classification accuracies range from 86% to 78% for the water and land classifiers, with accuracies at 88% or greater for general height-dependent groupings. Details of the databases, feature selection method, and classifiers, as well as example simulations, are presented.

### 1. Introduction

The proliferation of satellite imagery today is surpassing our manual ability to provide meaningful and timely analyses for the user. This statement is particularly true for the U.S. Navy, for which frequent tour rotations preclude the buildup of imagery interpretation expertise at remote, worldwide locations, particularly aboard ship. For these reasons, the Naval Research Laboratory (NRL) has, over the past 8–10 years, applied pattern recognition technology, now commonly called computer vision, to meteorological satellite imagery. This research has focused on cloud system (Peak and Tag 1992, 1994; Bankert and Aha 1995), as well as individual cloud, classification (Bankert 1994; Bankert and Aha 1996). Recently, automated tropical cyclone intensity analysis has been undertaken (Bankert and Tag 1998).

A number of researchers have addressed cloud classification from a variety of perspectives. Other imagery classification studies include, but are not limited to, the discrimination of surface and cloud types in polar regions (Ebert 1987, 1989; Key et al. 1989; Key 1990; Welch et al. 1992), the classification of cloud types in tropical scenes (Shenk et al. 1976; Desbois et al. 1982; Inoue 1987), the discrimination of ice and water clouds

(Knottenberg and Raschke 1982), the separation of clouds and snow (Tsonis 1984; Allen et al. 1990), and the classification of ocean clouds (Garand 1988). Lee et al. (1990) concluded that nonparametric classification methods such as neural networks and  $K$  nearest neighbor produced results superior to, and needed less training data than, linear discriminant analysis in the classification of stratocumulus, cumulus, and cirrus. More recently, Lewis et al. (1997) added spatial and temporal cloud characteristics to conventional spectral and textural cloud features in an attempt to improve classification accuracy. Using Advanced Very High Resolution Radiometer (AVHRR) imagery, Miller and Emery (1997) developed an automated classifier for both underlying land and ocean surfaces; the technique is intended for determining the choice of a passive microwave rain-rate algorithm for use with the Tropical Rainfall Measuring Mission satellite. Baum et al. (1997) applied fuzzy logic in their classification approach to discriminate between single- and multilayered clouds. Lubin and Morrow (1998) applied Ebert's (1987) classifier to overwater AVHRR Arctic scenes in which surface observations of cloudiness were used as ground-truth verification.

Cloud classification procedures by the above authors serve diverse purposes. Our purpose is to develop a classifier that can provide a quick "first-look" analysis of conventional meteorological cloud types to a naval meteorologist. Because the only digital imagery data that U.S. Navy ships can acquire directly is from orbiting satellites, we limited this initial work to AVHRR

---

*Corresponding author address:* Dr. Paul M. Tag, Department of the Navy, Naval Research Lab, Monterey, CA 93943-5502.  
E-mail: tag@nrlmry.navy.mil

TABLE 1. Geographic regions and months of the year (first J is Jan, etc.) used for water cloud database (AVHRR imagery).

	J	F	M	A	M	J	J	A	S	O	N	D
China Sea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Western Atlantic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Persian Gulf		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Mediterranean Sea				✓	✓	✓			✓		✓	
Kuroshio Current			✓	✓	✓	✓	✓		✓	✓		
Pacific Ocean	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
Barents Sea				✓	✓	✓	✓	✓				

[in contrast to Defense Meteorological Satellite Program (DMSP)] imagery because of its availability. As with other artificial intelligence (AI) products that we have developed, such as expert systems, we treat this product as an aid to the forecaster and not an end in itself. Note that, because we originally expected to apply any developed algorithm to DMSP data, which has only corresponding visible and infrared channels, the AVHRR development was limited to the use of channels 1 and 4 only. Current development of a Geostationary Operational Environmental Satellite (GOES)-based classifier will consider additional channels.

Section 2 discusses the development of the cloud databases used for our supervised cloud classifications. Section 3 covers the feature selection and classification procedures. Section 4 summarizes the classification accuracies for the water/land daytime components of the classification system. Section 5 discusses postprocessing and display. Section 6 provides a summary of lessons learned and final discussion.

## 2. Cloud databases

### a. Requirements/maritime database

In general, classification procedures can be divided into two types: supervised and unsupervised. The maritime daytime classification procedure developed by Bankert (1994) was based upon supervised classification, using a database of NRL-classified cloud images (Crosiar 1993), supplemented by approximately 600 classifications from the Naval Postgraduate School (C. Wash 1991, personal communication). See Table 1 for a listing of the regions and months covered by the maritime database.

The premise of supervised classification is the “training” of a classifier based upon known, typed cases of specific clouds (in a  $16 \times 16$  pixel area for our purposes) such that the classifier, once trained, can be used with confidence on unknown cloud image samples. This method, although straightforward, entails considerable effort in the manual typing of the training samples. Bankert’s (1994) database had a total of 1633 samples in 10 classes [later increased to 11 classes (see Table 2), for a total of 1834 samples, to allow for the addition of cumulus congestus (CuC) and altocumulus (Ac), and the removal of nimbostratus (Ns)]. Manual classifications by the experts included the following requirements: stratiform clouds had to cover 75% or greater of a  $16 \times 16$  pixel area; for cumuliform clouds (except Cb), the coverage requirement was 50%. Note that the only cirrus-type cloud allowed to be defined as optically thin was Ci (not Cs or Cc). Samples did not include any areas covered by smoke or fires. Considering the multitude of these complexities, it would be very desirable if training data could be generated without the time-consuming process of manual classification.

One automated method for creating such a database would be to use unsupervised learning techniques. An unsupervised method allows the classifier to determine, using a mathematical separability of classes based upon the designated cloud characteristics (features), its own division of cloud types. These types may or may not coincide with conventional typing nomenclatures, but are often useful in understanding new sets of data for which there are no recognized separations. In an attempt at determining whether unsupervised classification could be used to generate a land database, Gordon et al. (1995) applied two unsupervised learning procedures, the K-Means algorithm (MacQueen 1967) and AutoClass (Cheeseman et al. 1988), to our previously hand-classified maritime cloud database. Our goal was to determine if existing clustering algorithms could type clouds into classes similar to conventional meteorological cloud classes (Table 2). Without presenting the detailed results here, we concluded that the classes developed by the clustering methods, although generally consistent with the classes shown in Table 2, were not sufficiently segregated in order to formulate a new land database.

TABLE 2. Cloud classes used for typing cloud imagery.

Cirrus (Ci)
Cirrocumulus (Cc)
Cirrostratus (Cs)
Altostratus (As)
Altostratus (As)
Stratocumulus (Sc)
Stratus (St)
Cumulus (Cu)
Cumulus Congestus (CuC)
Cumulonimbus (Cb)
Clear (Clr)

TABLE 3. Geographic regions/months/sample number used for overland cloud database.

	J	F	M	A	M	J	J	A	S	O	N	D
Eastern U.S./northeastern Canada	9		39	32				14				
Eastern Med./Europe and northeastern Africa	21	11		11	17				13		21	
Western Med./Europe and northwestern Africa		18	20		6		13		22	40	12	
Central Med./Europe and North Africa	4		15	10			14		12			10
Eastern China				96				69	7			
Northwest Norway/Sweden/Finland		23	66	55	67	18	29	47	25			
Soviet Union					56		14					
Northeast Soviet Union						10	15					
Northwest U.S./southwest Canada	11							17				17
Southwest U.S.				84				42				
Iraq/Saudi Arabia/Iran			14		66		98	65	62	13	87	10
Greenland/Iceland, U.K.		22								13		
Northwest Africa										3		

b. Land database

Because of limitations and uncertainties imposed by using an unsupervised learning procedure, the general procedure established by Crosiar (1993) was followed again for development of the cloud database over land. Over a 2-month period during the spring of 1996, three meteorological satellite experts (Mr. Robert Fett and Mr. Ron Englebretson of Science Applications International Corporation, and Mr. Robin Brody of NRL) were employed to hand-classify cloud samples on pixel data taken from AVHRR local area coverage images, the same as used by Crosiar. The AVHRR instrument collects radiance measurements of reflected solar radiation and emitted thermal infrared (IR) energy in five spectral window regions at a spatial resolution of 1.1 km at nadir. Identical to Bankert's (1994) scheme, visible channel 1 (0.55–0.68 μm) and IR channel 4 (10.5–11.5 μm) were chosen as the channels to be used in the classification procedures. As was done with the maritime cloud data, visible channel data are normalized by the cosine of the solar zenith angle (SZA). For both the training data and later operational application, an SZA of less than 75° was arbitrarily defined.

Data from 13 worldwide regions were used for the training/testing dataset; see Table 3 for a listing. In addition to the visible and IR imagery, the human classifiers had pixel-level temperatures (computed from the IR data) and surface and upper-level synoptic charts at their disposal. From the imagery represented in Table

3, a total of 7912 16 × 16 (see section 3 for size restrictions) pixel cloud samples were analyzed by the three experts for the cloud classes shown in Table 2. Of these, all experts agreed on the classes of 2945 samples. However, because many of these agreements were for samples that fell into an unknown or “mixed” cloud category, only 1201 samples were of use in a supervised classifier. Samples were considered mixed if more than one cloud type was determined to exist in the 16 × 16 pixel region. No mixed samples are used in the training data. With 7 of 11 classes falling below the nominal 100 samples per class (see section 2a), we decided to augment this number with an additional 421 samples in which two out of three experts agreed, and the third differed only in type but not in height of cloud (low, middle, high). For example, if two experts classified an image as Sc and the third as Cu, the Sc classification was accepted; however, if two classified an image as As and the third registered Cs, this sample was thrown out (see Table 4).

Eliminating an additional 88 samples for either noise or SZA restrictions, and later adding 71 independent samples for cirrocumulus (Cc, for which there were only seven agreed-upon samples in the original database), produced a total of 1605 samples, distributed as shown in Table 5. Note that there are disproportionately more samples in the clear category. However, the number of clear samples was not limited for two reasons: these

TABLE 4. Agreement of three experts in the classification of 7912 samples in the land database.

	Entire dataset	Entire dataset minus “mixed”
Three out of three	2945	1201
Two out of three	3747	2062
Two out of three, with no “height” disagreement	—	727
No agreement	1220	—
Total	7912	

TABLE 5. Cloud classes and number of samples used for typing land imagery.

Cirrus (Ci)	199
Cirrostratus (Cs)	116
Cirrocumulus (Cc)	78
Altostratus (As)	101
Cumulus Congestus (CuC)	110
Stratocumulus (Sc)	162
Stratus (St)	105
Cumulus (Cu)	182
Cumulonimbus (Cb)	90
Clear (Clr)	372
Total	1605

TABLE 6. Separation of total feature subset into textural, spectral, and physical definitions.

	Overwater	Overland
Textural	86	160
Spectral	14	14
Physical	10	7
Total	110	181

samples had agreement from all three experts and represent an important component for the classifier—a worldwide database of land backgrounds valid nearly year-round. Except for this qualitative worldwide sampling, no sampling distinction was made with regard to vegetation, soil type, or wetness.

### 3. Feature selection and classification

The visible and IR imagery databases form the basis for computing features used to classify  $16 \times 16$  pixel areas into the 11 cloud classes shown in Table 2. We separate features, for both water and land classifications, into one of three categories: spectral (e.g., maximum and minimum pixel values), textural (spatial distribution of pixel values), and physical (e.g., cloud fraction, latitude). Because textural features are computed within subregions ( $4 \times 4$  pixel areas) of the samples, a  $16 \times 16$  pixel area is the minimum area for which textural characteristics can be defined. For the results presented here, there are 181 total features for the land classifications, and 110 for the water, calculated from this  $16 \times 16$  pixel area (see Table 6). [The original water classifications described by Bankert (1994) used over 200 features; for technical reasons, this total feature set was reduced with no loss in accuracy.]

From this large number of features, we must isolate

those features that optimize classification accuracy. Because there is redundant or irrelevant information within the feature set, the use of all features necessarily degrades classifier performance. Feature reduction is performed with a feature selection algorithm developed by Dr. David Aha from the NRL AI Center in Washington, District of Columbia. This algorithm, which uses backward sequential selection (BSS), is embedded with a 1-nearest neighbor (1-NN) classifier acting as the evaluation function. In effect, different randomly selected feature subsets are presented to the algorithm. Within the algorithm, BSS is a process in which features are progressively removed from the feature subset until performance (using 1-NN evaluation) does not improve; see Aha and Bankert (1995) and Bankert and Aha (1996). The reduced subset that provides the highest accuracy in a cross-validation test is selected. The feature selection sequence is illustrated in Fig. 1.

The feature selection testing procedure uses leave-one-out (LOO) cross validation of the 1-NN classifier. A LOO test is an extreme version of cross validation. Whereas, for example, 10-fold cross validation removes 10% of the dataset each iteration to be used as testing data (with the remaining 90% used for training), with this process repeated 10 times, a LOO cross validation repeats the training and testing the same number of times as there are data samples; if there were 1000 data samples, the training and testing process would be repeated 1000 times, with each sample serving as an independent test based on a training set of 999. The average of the 1000 tests would provide an estimate of the algorithm's accuracy on unseen data.

In operations, the 1-NN classifier is used with the feature set selected by the BSS/1-NN algorithm. This procedure is different from that of Bankert (1994), who used a probabilistic neural network (PNN) to do the

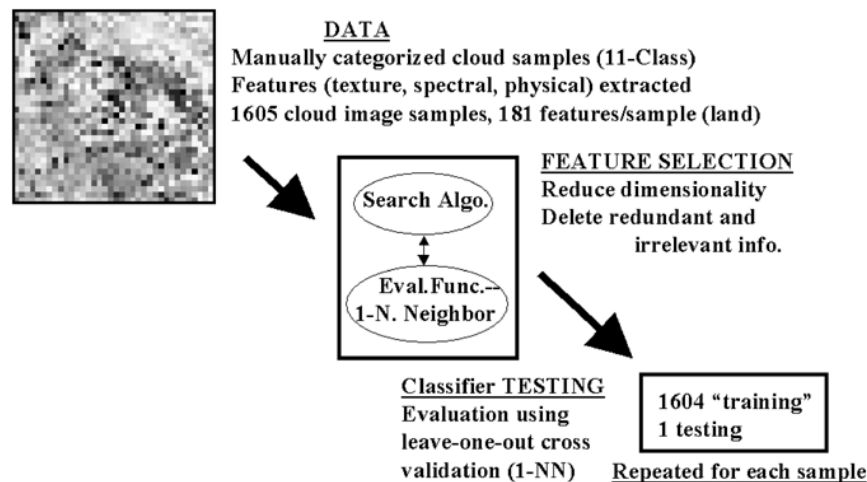


FIG. 1. Feature selection procedure. From the original set of features, a subset is selected and tested to define an overall classification accuracy. This process is repeated over and over, with different feature subsets, to iterate to that set of features that optimizes overall 11-class accuracy in a leave-one-out test.

TABLE 7. Selected water features.

	Channel
Spectral	
Pixel value maximum	Visible
Pixel value median	Visible
Pixel value min (max temp)	IR
Pixel value median	IR
Pixel value mean	IR
Textural	
GLDV mean difference (4 × 4 mean)	Visible
GLDV mean difference (16 × 16)	IR
GLDV mean difference (4 × 4 min)	IR
GLDV cluster shade (16 × 16)	Visible
Physical	
Latitude	—

TABLE 8. Selected land features.

	Channel
Spectral	
Pixel value max (min temp)	IR
Pixel value mean	IR
Textural	
GLDV SD	Visible
SADH entropy	IR
SADH SD	Visible
SADH SD (max)	Visible
SADH mean sum	Visible
SADH mean sum (max)	IR
SADH mean sum (min)	IR
Physical	
Cloud-top temperature	
Month—ABS (month 6.5)	—

classification. The changeover to a 1–NN was made because of its use in the feature selector. Because the PNN requires manual adjustment of a smoothing parameter to optimize performance, the PNN was not suitable for the iterative feature selection process. Features selected using the 1–NN as the evaluation function were usually not ideal, with some accuracy lost, when used as the PNN input vector. As noted by Lee et al. (1990), both the 1–NN and the PNN are nonparametric classifiers, referring to the fact that no a priori assumption is made regarding the statistical distributions of the supporting feature sets. This characteristic permits accuracies for the 1–NN and the PNN to be similarly high.

From the initial feature sets summarized in Table 6, and using the Aha feature selection program, 10 features for water and 11 features for land provided the highest 1–NN LOO accuracy. See Tables 7 and 8. Textural measures, representing the spatial distribution of pixel gray-level values within the image, are defined as GLDV, referring to gray-level difference vector, and SADH, referring to sum and difference histogram computations. ASM is the angular second moment; SD, standard deviation; and SZA, solar zenith angle; ABS refers to

absolute value. See Welch et al. (1992) for a detailed description of texture measurements.

#### 4. Accuracy

LOO testing, as described in section 3, is used to compute accuracies. As noted earlier, redundant or irrelevant information within the feature set degrades classifier performance—for example, the use of all features (Table 6) produces an 11-class accuracy of 73.5% for water and 68.0% for land. Specific portions of the feature database produce lower accuracies as well—the use of spectral features alone results in 76.2% and 68.0% for water and land, respectively. Using a mixture of feature types almost always produces a higher accuracy. Detailed accuracies using the selected features listed in Tables 7 and 8 are now presented.

##### a. 11-class

Tables 9 and 10 provide LOO testing accuracies for the water and land databases, respectively. The overall,

TABLE 9. Confusion matrix (leave-one-out test) for water database (overall accuracy—86.4%). This table shows relative accuracies, based on the training data, for each of the cloud classes. Table rows represent actual classes and columns represent the predicted. As an illustration, consider the row for Cirrus (Ci). Of the 212 training cases, 188 (89%) were analyzed correctly, with 1 case misclassified as Cirrocumulus (Cc), 15 misclassified as Cirrostratus (Cs), etc.

Cloud	Ci	Cc	Cs	Ac	As	Sc	St	Cu	CuC	Cb	Cl	%	No.
Ci	<b>188</b>	1	15		3	1	2			2		<b>89</b>	212
Cc	1	<b>66</b>	2		1					2		<b>92</b>	72
Cs	29	3	<b>124</b>		4					5	1	<b>75</b>	166
Ac	1			<b>88</b>	13	8	6	1	6			<b>72</b>	123
As	2	2	1	7	<b>131</b>	2	3		1	4	1	<b>85</b>	154
Sc	2			4	2	<b>217</b>	6	11	8	1		<b>87</b>	251
St	1		2	4	4	13	<b>197</b>	2	1		1	<b>88</b>	225
Cu						8	4	<b>107</b>	4			<b>87</b>	123
CuC	1			3	2	9	2	3	<b>116</b>		1	<b>85</b>	137
Cb	3	3	5	1	4					<b>133</b>		<b>89</b>	149
Cl	1					2					<b>218</b>	<b>98</b>	222
Total													1834

TABLE 10. Same as Table 9, but for land database (overall accuracy—77.5%).

Cloud	Ci	Cc	Cs	Ac	As	Sc	St	Cu	CuC	Cb	Cl	%	No.
Ci	<b>147</b>	26	12	1	4					9	2	<b>74</b>	199
Cc	21	<b>45</b>	7							5		<b>58</b>	78
Cs	14	13	<b>87</b>		1					1		<b>75</b>	116
Ac	7		0	<b>51</b>	14	14		2	2			<b>57</b>	90
As	5	2	1	21	<b>57</b>	9	3		1	1		<b>56</b>	101
Sc				7	8	<b>121</b>	9	10	4		3	<b>75</b>	162
St				1	2	13	<b>84</b>				4	<b>80</b>	105
Cu	1			1		3	1	<b>147</b>	24		5	<b>81</b>	182
CuC				1	0	5	1	23	<b>79</b>	1		<b>72</b>	110
Cb	8	7	3						5	<b>67</b>		<b>74</b>	90
Cl					2	4	3	4			<b>359</b>	<b>96</b>	372
Total													1605

11-class accuracies for the water and land classifiers are 86% and 78%, respectively. Individual cloud class percentages vary from 98 (Cl) to 72 (Ac) for water and 96 (Cl) to 57 (As) for land classifications. Overall, the water classifications are superior to those for land, probably because of the large disparity of land backgrounds in comparison with the relatively minor variations in underlying water conditions.<sup>1</sup> The lowest accuracy percentages for water occur for Cs (75) and Ac (72), with all other accuracies at 85% or higher. Over land the accuracies are not as good, with As (57%), Ac (57), and Cc (58) being the lowest, and other class accuracies above 70%. Although the relatively low accuracies of Ac and As are due partially to misclassifications between the two classes (see section 4b), many of the additional misclassifications are with the lower Sc. Also, notice the number of Cb samples misclassified as high clouds (Ci, Cc, and Cs) and vice versa. These latter misclassifications remind us that the algorithm is basing

its classifications on cloud-top information only, and that such a viewpoint by itself can lead to ambiguity. Inclusion of features computed from channel 3 (3.7  $\mu$ ) could have improved these accuracies and will be a part of the upcoming GOES classification project (see section 6).

#### b. 5-class

A close look at the confusion matrices (Tables 9 and 10) reveals that many of the misclassifications are between clouds of similar height (e.g., cirrus versus cirrostratus; altocumulus versus altostratus, etc.). For many users, a classification based primarily upon height provides a sufficient depiction without the meteorological detail provided by Latin-based cloud types. For this reason, the classifications have also been grouped into classes labeled low, middle, high, vertical, and clear. The “vertical” class was chosen to represent cumulus congestus and cumulonimbus, two classes that indicate convection of some significance. Tables 11 and 12 summarize these accuracies, which, by definition, are higher, with average overall values of 92.7% and 88.1%, respectively, for the water and land classifications. Simple groupings of 11 classes into 5 classes were shown to have approximately the same accuracy as selecting features based upon a 5-class dataset (as done in Bankert 1994). The lowest accuracy over water is 86%, for middle-level clouds. Over land, the lowest accuracy is also for middle clouds (75%), with vertical clouds at 76%.

<sup>1</sup> We realize that region-specific classifiers would produce better results, but our limited numbers of samples would make such a classifier less robust. However, the distribution of the misclassifications leads us to believe that this single classifier is acceptable. Application to regional data on a worldwide basis will provide optimum testing.

TABLE 11. Five-class confusion matrix (leave-one-out test) for water database (overall accuracy—92.7%). This table shows relative accuracies, based on the training data, for each of the cloud groupings: low, middle, high, vertical, and clear. Table rows represent actual classes and columns represent the predicted. As an illustration, consider the row for low clouds. Of the 599 training cases, 565 (94%) were analyzed correctly, with 14 misclassified as middle clouds, 5 as high clouds, etc.

Cloud grouping	Low	Middle	High	Vertical	Clear	%	No.
Low	<b>565</b>	14	5	14	1	<b>94</b>	599
Middle	20	<b>239</b>	6	11	1	<b>86</b>	277
High	3	8	<b>429</b>	9	1	<b>95</b>	450
Vertical	14	10	12	<b>249</b>	1	<b>87</b>	286
Clear	2		1	1	<b>218</b>	<b>98</b>	222
Total							1834

## 5. Processing and display

The classifications described in section 4, based upon the training data, consider individual, hand-typed samples of 16  $\times$  16 pixel areas. When the samples are not as pure as the training set (i.e., mixed cloud types), our “winner-take-all” classifiers type every 16  $\times$  16 pixel area into a single cloud type. In practice, this classification procedure must be applied sequentially to an entire image. Rather than moving 16 pixels at a time across the grid (resulting in the minimum amount of computation), the classifier is applied every 8 pixels instead—

TABLE 12. Same as Table 11, but for land database (overall accuracy—88.1%).

Cloud group- ing	Low	Middle	High	Vertical	Clear	%	No.
Low	<b>388</b>	19	1	29	12	<b>86</b>	449
Middle	28	<b>143</b>	15	4	1	<b>75</b>	191
High		6	<b>372</b>	15		<b>95</b>	393
Vertical	29	1	18	<b>152</b>		<b>76</b>	200
Clear	11	2		1	<b>359</b>	<b>97</b>	372
Total							1605

see Fig. 2. As a result, most pixels (except those near the edge of the domain) are classified four times rather than once. The dominant classification for a pixel is the chosen classification. This method, although taking more computational time, provides more resolution and a softer appearance for cloud boundaries. Dependent upon the underlying surface, the water or land classifier is used to classify each  $16 \times 16$  pixel area in an unregistered image. Along a coast line, the choice of classifier is dependent upon where (land or water) the majority of pixels in a square reside.

The classifications and corresponding satellite images are available for viewing on the satellite page of the NRL Monterey Web site ([www.nrlmry.navy.mil/projects/sat\\_products.html](http://www.nrlmry.navy.mil/projects/sat_products.html)) on the Internet. Sample output from the classifiers, illustrating typical winter and summer West Coast weather patterns, is shown in Figs. 3 and 4.<sup>2</sup> The Web page allows several display options. Both unregistered and registered satellite imagery (with both visual and IR) and classifications are available. For unregistered imagery, the entire satellite pass is shown. For the registered image, a specific West Coast subsection is isolated and redisplayed in Mercator coordinates. (Note that for unregistered imagery from a satellite perspective, all pixels are the same size; in transforming to a Mercator projection for the registered imagery, pixel size becomes distorted.) For the Web presentations, both 11- and 5-class displays are available. Cloud types are color-coded, with general color classifications as follows: yellow, high clouds; green, midlevel; blue, low level; and red, vertical (CuC and Cb).

Evaluating classification output and making improvements have been ongoing since the summer of 1997. One of the obvious observations was that single clas-

<sup>2</sup> For the winter depiction shown in Fig. 3, the classifier distinguishes the various frontal clouds (cirrostratus/cirrus tapering off to altocumulus/altostratus toward the southwest where the front weakens) with embedded Cb associated with the frontal band. Also shown, as one would expect, are the cumulus congestus that predominate behind the front on the cyclonic side of the polar jet stream, while stratocumulus predominates ahead of the front on the anticyclonic side of the polar jet. For the summer case (Fig. 4), the classifier shows the typical stratocumulus found within the boundary layer over the eastern north Pacific. The middle and high clouds in the northwest corner are associated with an approaching trough.

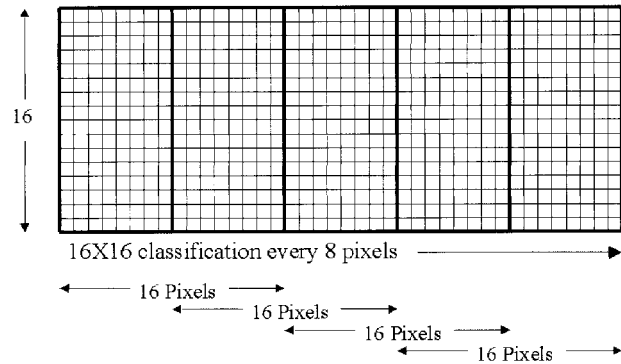


FIG. 2. Cloud classification is performed over a  $16 \times 16$  unregistered pixel area. The classifier is applied every eight pixels across the domain (both N-S and E-W), with all pixels (except near the edge) being classified nine times. The dominant classification for a pixel is the chosen classification.

sifications for each  $16 \times 16$  pixel area visually dominate regions for which we know that clouds are not continuous, particularly those involving cumuliform clouds such as cumulus, stratocumulus, cumulus congestus, and cumulonimbus. For these cloud types, there can be significant clear areas between the clouds, particularly at the AVHRR resolution of 1 km. For this reason, one of our colleagues (J. E. Peak 1997, personal communication) made the suggestion that a postprocessing step might be in order: after all classifications are performed for the entire image, we examine each image pixel to determine if that pixel represents clear conditions. This process, although more complicated over land than over water, can be a simple test based upon whether the pixel albedo (visible channel) lies below a certain threshold. For example, we found that an albedo of 15% (after correcting for SZA) provides a conservative threshold. If a pixel is deemed to be clear, that pixel is then set to black, representing clear conditions. This postprocessing step, along with the computational smoothing technique discussed earlier, has the significant benefit of making the classification pattern appear, visually, much more realistic. Future work may involve enhancing this postprocessing step to include pixel temperature as well.

## 6. Discussion and lessons learned

The cloud classification algorithms used to produce the displayed classifications over land and water (see Figs. 3 and 4) have been evaluated over the eastern Pacific and the western and central United States during the summer, fall, and winter months using NOAA-14 imagery valid between 2000 and 2300 UTC. During recent months we have evaluated morning NOAA-12 and NOAA-15 imagery as well. Some classifier strengths have become evident. For example, aside from generally acceptable classifications of small cloud groupings, the distribution of cloud types around midlatitude storm systems appears reasonable. The postprocessing step in

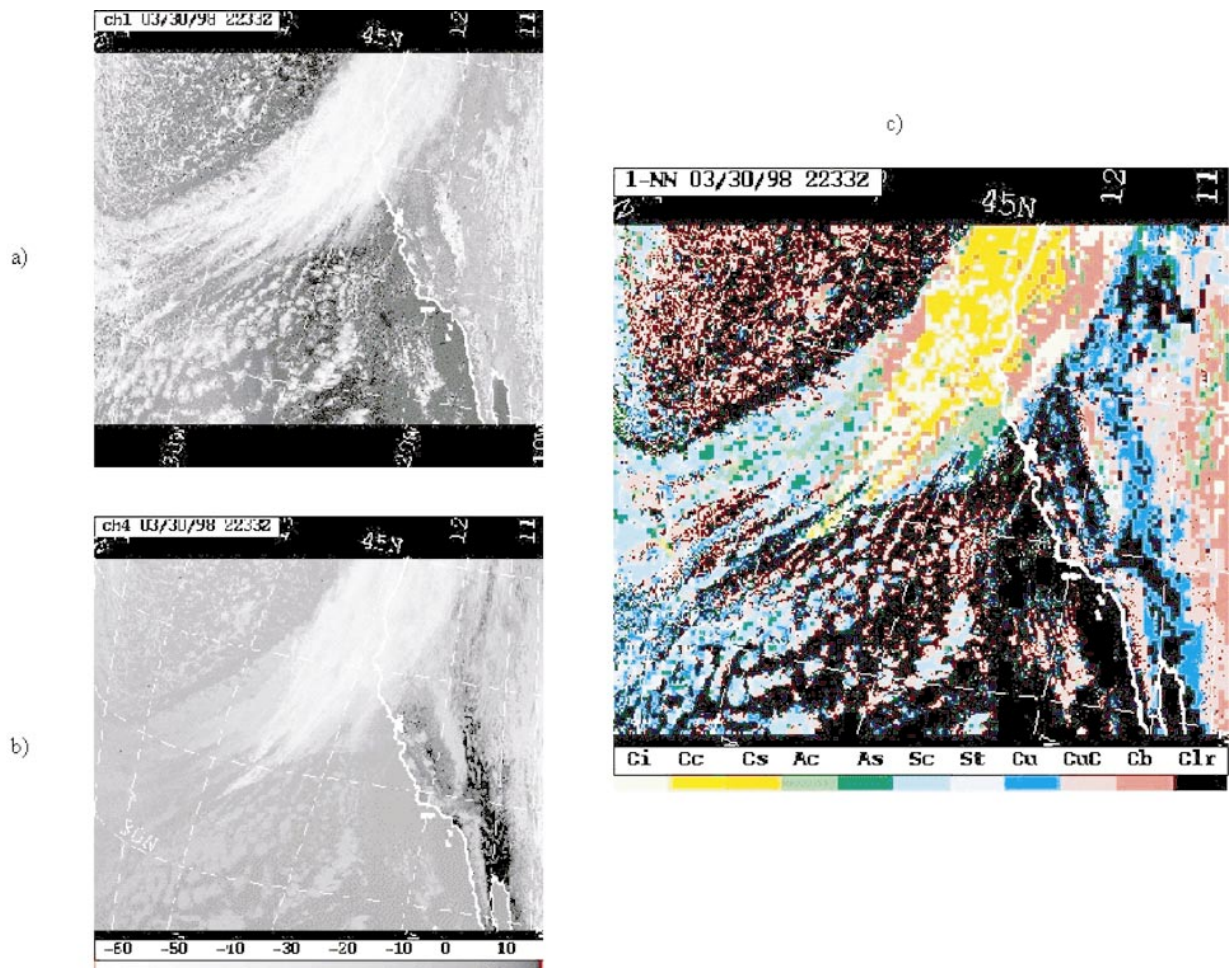


FIG. 3. Wintertime frontal band approaching the U.S. West Coast: (a) visible image, (b) infrared image, (c) 11-class cloud classification.

which a pixel-by-pixel examination identifies clear pixels within classified ( $16 \times 16$  pixel) cloud areas results in a more accurate product, particularly for cumuliform clouds.

During early development, to maximize LOO accuracy, we included region as a feature for the land classifier. Both the original 12 regions and 4 integrated, more general categories—desert, high latitude polar, eastern continental (such as east Asia and eastern North America), and western continental (including the Mediterranean)—were tried. While including region resulted in a several percent gain in LOO accuracy, testing revealed, in practice, occasional geographic classification discontinuities when more than one region existed in a satellite pass. This result stems primarily from the fact that the training database is far from being complete in terms of cloud types for all regions. For this reason, region was eliminated as a feature. We understand that this limitation means that the classifier must be retested for any new region in which the classifier is applied.

As with any numerical analysis based upon computer vision, the product must be interpreted and its idiosyn-

crasies noted. The following list summarizes a few of the classifier's weaknesses.

- Because entirely different classifiers (in terms of features and training data) are used on either side of a coastline, an occasional discontinuity in cloud type can occur across that boundary.
- For the few cases that we observed last summer in which sunglint was seen over the water, the classifier appeared not to be fooled by the increased reflectivity. However, recent classifications of high-angle *NOAA-12* morning imagery show that sunglint can be labeled as a low cloud. As a result, sunglint areas must be interpreted with caution. Because the location of sunglint can be determined mathematically, one option being considered for presentation purposes is an automated delineation of sunglint areas.
- The classifier can be fooled by snow-covered surfaces, particularly over high terrain. This weakness has been noted in springtime classifications for the Sierra Nevada of California. Fortunately, an analyst can discern this anomaly because of the recurring patterns from day to day.



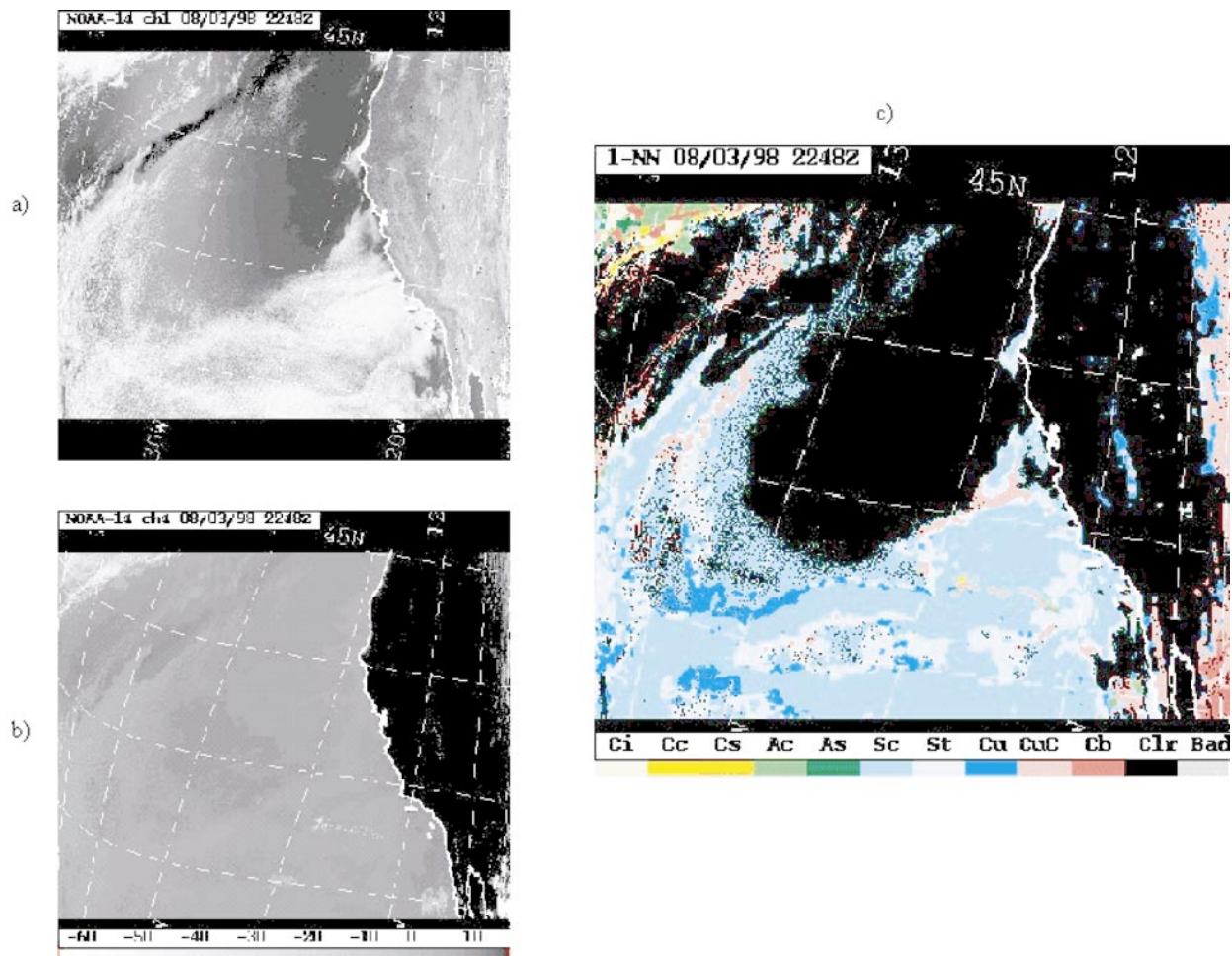


FIG. 4. Same as Fig. 3, except for low-cloud pattern typical of summertime eastern Pacific.

- Classifications over water tend to overestimate the occurrence of Cb.

Our near-term plans involve evaluating the classifiers in other parts of the world, the first two being the eastern United States and the European/Mediterranean area, geographic regions covered by the U.S. Navy's regional weather centers in Norfolk, Virginia, and Rota, Spain. In addition, we are developing software to create expert-defined databases for a similar GOES classifier. Development of this classifier should proceed more smoothly, evolving from our experience thus far, and produce changes that address many of the weaknesses noted above. Improvements, such as including the 3.9- $\mu$  GOES channel, should provide a more robust, accurate, and useful classifier that can take advantage of the geostationary satellite's high temporal frequency.

*Acknowledgments.* The support of the sponsors, the Office of Naval Research under Program Element 0602435N, and the Oceanographer of the Navy through

the program office at the Navy's Space and Naval Warfare System Command under Program Element 0603207N, is gratefully acknowledged. The authors wish to thank the following individuals: Mr. Robert W. Fett and Mr. Ron Englebretson of Science Applications International Corp. (SAIC) for their careful typing of the 8000 land cloud samples; Dr. David Aha of NRL for his feature selection algorithm; Mr. Jonathan Allen of Computer Sciences Corporation (CSC) for implementing the classifier code and developing the classification graphical display; Mr. John Kent, of SAIC, for the overall Web page design and the integration of the cloud classifications into the page; and Mr. James E. Peak of CSC for writing the summarization code for the land database and his suggestions for an improved classification display.

REFERENCES

Aha, D. W., and R. L. Bankert, 1995: A comparative evaluation of sequential feature selection algorithms. *Learning from Data: Ar-*

- tificial Intelligence and Statistics V*, D. Fisher and H.-J. Lenz, Eds., Springer-Verlag, 199–206.
- Allen, R. C., Jr., P. A. Durkee, and C. H. Wash, 1990: Snow/cloud discrimination with multispectral satellite measurements. *J. Appl. Meteor.*, **29**, 994–1004.
- Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909–918.
- , and D. W. Aha, 1995: Automated identification of cloud patterns in satellite imagery. Preprints, *14th Conf. on Weather Analysis and Forecasting*, Dallas, TX, Amer. Meteor. Soc., 313–316.
- , and —, 1996: Improvement to a neural network cloud classifier. *J. Appl. Meteor.*, **35**, 2036–2039.
- , and P. M. Tag, 1998: Using SSM/I data and computer vision to estimate tropical cyclone intensity. Preprints, *Ninth Conf. on Satellite Meteorology and Oceanography*, Paris, France, Amer. Meteor. Soc., 226–229.
- Baum, B. A., V. Tovinkere, J. Titlow, and R. M. Welch, 1997: Automated cloud classification of global AVHRR data using a fuzzy logic approach. *J. Appl. Meteor.*, **36**, 1519–1540.
- Cheeseman, P., J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman, 1988: Autoclass: A Bayesian classification system. *Proc. Fifth Int. Conf. on Machine Learning*, Ann Arbor, MI, Cognitive Science and Machine Intelligence Laboratory of the University of Michigan, 54–63.
- Crosiar, C. L., 1993: An AVHRR cloud classification database typed by experts. Naval Research Laboratory Memo. Rep. NRL/MR/7531-93-7207, Monterey, CA, 31 pp. [Available from Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943-5502.]
- Desbois, M., G. Seze, and G. Szejwach, 1982: Automatic classification of clouds on Meteosat imagery: Application to high-level clouds. *J. Appl. Meteor.*, **21**, 401–412.
- Ebert, E., 1987: A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *J. Climate Appl. Meteor.*, **26**, 1412–1427.
- , 1989: Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme. *J. Appl. Meteor.*, **28**, 382–399.
- Garand, L., 1988: Automated recognition of oceanic cloud patterns. Part I: Methodology and application to cloud climatology. *J. Climate*, **1**, 20–39.
- Gordon, D. F., P. M. Tag, and R. L. Bankert, 1995: Unsupervised classification procedures applied to satellite cloud data. Naval Research Laboratory Internal Rep. AIC-95-005, Washington, D.C., 6 pp. [Available from Naval Research Laboratory, Code 5514, Washington, DC 20375-5337.]
- Inoue, T., 1987: A cloud type classification with NOAA-7 split-window measurements. *J. Geophys. Res.*, **92**, 3991–4000.
- Key, J., 1990: Cloud cover analysis with arctic AVHRR data. Part II: Classification with spectral and textural measures. *J. Geophys. Res.*, **95**, 7661–7675.
- , J. A. Maslanik, and A. J. Schweiger, 1989: Classification of merged AVHRR and SMMR arctic data with neural networks. *Photogramm. Eng. Remote Sens.*, **55**, 1331–1338.
- Knottenberg, H., and E. Raschke, 1982: On the discrimination of water and ice clouds in multispectral AVHRR data. *Ann. Meteor.*, **18**, 145–147.
- Lee, J., R. C. Weger, S. K. Sengupta, and R. M. Welch, 1990: A neural network approach to cloud classification. *IEEE Trans. Geosci. Remote Sens.*, **28**, 846–855.
- Lewis, H. G., S. Cote, and A. R. L. Tatnall, 1997: Determination of spatial and temporal characteristics as an aid to neural network cloud classification. *Int. J. Remote Sens.*, **18**, 899–915.
- Lubin, D., and E. Morrow, 1998: Evaluation of an AVHRR cloud detection and classification method over the central Arctic Ocean. *J. Appl. Meteor.*, **37**, 166–183.
- MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, L. M. LeCam and J. Neyman, Eds., University of California Press, 281–297.
- Miller, S. W., and W. J. Emery, 1997: An automated neural network cloud classifier for use over land and ocean surfaces. *J. Appl. Meteor.*, **36**, 1346–1362.
- Peak, J. E., and P. M. Tag, 1992: Toward automated interpretation of satellite imagery for Navy shipboard applications. *Bull. Amer. Meteor. Soc.*, **73**, 995–1008.
- , and —, 1994: Segmentation of satellite imagery using hierarchical thresholding and neural networks. *J. Appl. Meteor.*, **33**, 605–616.
- Shenk, W. E., R. J. Holub, and R. A. Neff, 1976: A multispectral cloud type identification method developed for tropical ocean areas with Nimbus-3 MRIR measurements. *Mon. Wea. Rev.*, **104**, 284–291.
- Tsonis, A. A., 1984: On the separability of various classes from the GOES visible and infrared data. *J. Climate Appl. Meteor.*, **23**, 1393–1410.
- Welch, R. M., S. K. Sengupta, A. K. Goroch, P. Rabindra, N. Rangaraj, and M. S. Navar, 1992: Polar cloud and surface classification using AVHRR imagery: An intercomparison of methods. *J. Appl. Meteor.*, **31**, 405–420.