Classification of Merged AVHRR and SMMR Arctic Data with Neural Networks

J. Key, J. A. Maslanik, and *A. J. Schweiger*
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309–0449

ABSTRACT: A forward-feed back-propagation neural network is used to classify merged AVHRR and SMMR summer Arctic data. Four surface and eight cloud classes are identified. Partial memberships of each pixel to each class are examined for spectral ambiguities. Classification results are compared to manual interpretations and to those determined by a supervised maximum likelihood procedure. Results indicate that a neural network approach offers advantages in ease of use, interpretability, and utility for indistinct and time-variant spectral classes.

INTRODUCTION

THE ARCTIC REGION provides a unique set of problems for image analysis algorithms. Current procedures for automated analyses of satellite radiance data have been developed for low and middle latitudes, but their application to polar regions has been largely unexplored. Those that have been applied often fail in the polar regions because surface temperatures are commonly as low or lower than cloud-top temperatures, and because snow-covered surfaces exhibit albedos similar to those of clouds. Also, extremely low surface temperatures and solar illuminations cause satellite radiometers to operate near one limit of their performance range, and in winter no visible-wavelength data are available. Because of these problems, a complex analysis method is necessary (WMO, 1987). Classification of Arctic AVHRR data with clustering algorithms has been performed primarily by Ebert (1987, 1988), Key (1988), and Key *et al.* (1989).

In this paper, we investigate the ability of neural networks to extract four surface and eight cloud classes in the Arctic from a merged data set consisting of five Advanced Very High Resolution Radiometer (AVHRR) and two Scanning Multichannel Microwave Radiometer (SMMR) channels. Results are compared to manual interpretations and to output from a supervised classification with a maximum likelihood class assignment scheme. Because cloud and sea ice mapping for climatological studies requires the processing of many images covering large areas (for example, 30 days worth of AVHRR images for the northern hemisphere), selection of training sites in a supervised scheme or the assignment of spectral clusters to physical classes in an unsupervised approach can involve an unacceptable amount of time and effort. Because a class such as low cloud over ice actually includes a range of cloud thicknesses overlying a range of ice concentrations, considerable spectral variability exists within the class as well as within individual pixels. Our primary goal in this work is to investigate the ability of a neural network classifier to deal with the considerable within-class variability encountered in our data based on a relatively small training set. The manual and supervised classifications are used to provide benchmarks for comparison of the neural network results, rather than as a test of the merits of these more traditional methods.

Motivated by the apparent limitations of multispectral feature extraction from imagery and the availability of expert system development tools, artificial intelligence (AI) techniques have come into increased use for the analysis of remotely sensed data (e.g., Nicolin and Gabler, 1987; Matsuyama, 1987; McKeown, 1987; Estes *et al.*, 1986; Nandhakumar and Aggarwal, 1985; Campbell and Roelofs, 1984), and have also been employed in geographic information systems (GIS) applications (e.g., Usery and Altheide, 1988; Ripple and Ulshoefer, 1987; Robinson and Frank, 1987; Smith *et al.*, 1987; Jackson and Mason, 1986; Smith, 1984). Due to the limited knowledge of the physical processes in the environment and the inherent noise in many geophysical data, environmental systems often cannot be accurately represented through numeric values describing their physical properties and interactions, but rather are subjected to categorization into broad classes. Most applications of expert systems have sought to apply qualitative knowledge to decision-making; expert systems operate primarily on abstract symbolic structures. In remote sensing applications where pattern recognition, spatial and temporal context, and multivariate analysis are common requirements, coupled numeric/symbolic systems may be useful. This issue has recently been addressed by Kitzmiller and Kowalik (1987), Kowalik (1986), and Borchardt (1986). Traditional techniques may not be adequate to identify and make use of relationships across such a broad range of numeric and non-numeric variables.

The neural networks, or connectionist, approach was first introduced as a theoretical method of AI in the 1960s. However, limitations in simple systems were recognized by Minsky and Papert (1969) and the concept gave way to the symbol system approach for the next two decades. The idea has recently been revived due to advances in hardware technology allowing the simulation of neural networks and the development of nonlinear multi-layered architectures (Rumelhart *et al.*, 1986). The technique has considerable potential for remote sensing, as suggested by applications to automated pattern recognition (e.g., Ritter *et al.*, 1988). The relationship between symbolic AI and neural networks is addressed by Chandrasekaran *et al.* (1988).

DATA

The data sets used here provide a broad range of spectral information necessary to map clouds and surfaces in polar regions. These data are typical of the types of imagery used for mapping of global cloud, sea surface temperature, and other climatological variables. The Advanced Very High Resolution Radiometer (AVHRR) on board the NOAA-7 polar orbiting satellite measures radiance in five channels encompassing the visible, infrared, and thermal portions of the electromagnetic spectrum (1: 0.58 to 0.68μm, 2: 0.73 to 1.0μm, 3.55 to 3.93μm, 4: 10.3 to 11.3μm, 5: 11.5 to 12.5μm) with a nadir resolution of 1.1 km. Global Area Coverage (GAC) imagery is a reduced-resolution product created through on-board satellite processing, with each pixel representing a 3 by 5-km field of view (Schwalb, 1978). Channels 1 and 2 were converted to approximate spectral albedo; channels 3, 4, and 5 were converted to radiance in milliwatts/m^2-steradians-cm), then to brightness temperature (NOAA, 1984; Lauritsen *et al.*, 1979). The typically low water vapor content in the polar atmosphere and the low physical temperatures reduce most atmospheric effects to a point where

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 55, No. 9, September 1989, pp. 1331–1338

0099-1112/89/5509–1331$02.25/0

they may be neglected for the analyses performed here. Approximate corrections for solar zenith angle in channels 1 and 2 were accomplished through a division of the albedo by the cosine of the zenith angle.

The Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) is a conically scanning radiometer that senses emitted microwave radiation in five channels: 6.6, 10.7, 18.0, 21.0, and 37.0 GHz, with two polarizations (horizontal and vertical) per channel. At these frequencies, passive microwave data are relatively unaffected by clouds and provide useful data year-round independent of solar illumination. The 18 and 37 GHz vertical polarization channels are used here primarily for surface parameterization, with fields of view of 55 by 41 km and 27 by 18 km, respectively.

In order to study both clouds and surfaces beneath clouds, it is worthwhile to combine the AVHRR and SMMR channels into a single image set. AVHRR and SMMR data were merged in digital form and mapped to a polar sterographic projection. This projection yields pixels true to scale at 70° latitude with a five-kilometre pixel size. Five-kilometre pixels were averaged over 2 by 2 cells, yielding an effective pixel size of ten kilometres square. (Further constraints imposed by the image analysis system reduced this to 125 to 124 pixels.) SMMR data were converted to the five-kilometre cells by simple duplication of pixels. Further details are given in Maslanik *et al.* (1989). In this form, color composites can be made consisting of combinations of microwave, visible, and thermal-wavelength channels to highlight different cloud and surface features in the data.

The study area (Figure 1) is centered on the Kara and Barents Sea, extending north toward the pole and south toward Norway and the Siberian coast. Novaya Zemlya is near the center of the image. Shown are AVHRR channels 1, 3, and 4 for 1 July 1984. Both AVHRR and SMMR imagery were also acquired for 4 July 1984. While covering only a small portion of the Arctic Basin (1250 by 1250 km), it includes representative samples of all surface types found in the Arctic: snow-covered and snow-free land, sea ice of varying concentrations, and open water.

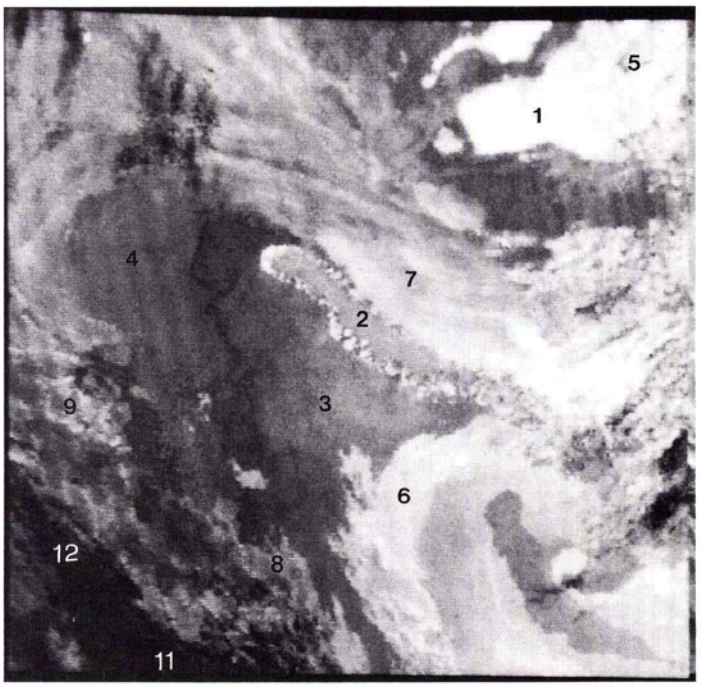

FIG. 1. The study area on 1 July 1984 showing Novaya Zemlya (at #2) in the center (75°N, 60°E), and the Kara (at #4 and #7) and Barents Seas (at #3). Sea ice covers most of the Kara Sea as well as the area north (toward lower left corner of image) and east (left) of Novaya Zemlya. Total area covered is approximately (1250 km)^2. Shown are AVHRR channels 1, 3, and 4. Examples of each target surface/cloud class are shown numbered as (1) LAND: snow-free land; (2) SNOW: snow-covered land/ice cap; (3) WATER: open water; (4) ICE: sea ice; (5) LCLL, (6) LCLW; (7) LCLI: low cloud over land, water, and ice, respectively; (8) MCLW; (9) MCLI: middle cloud over water, ice; (10) HCLL (now shown), (11) HCLW; and (12) HCLI: high cloud over land, water, ice. These classes were identified through a manual interpretation of AVHRR and SMMR data (see Figure 2).

METHODOLOGY

Four surface classes are of interest in this study: snow-free land, snow-covered land/ice cap, open water, and sea ice (abbreviations used: LAND, SNOW, WATER, and ICE). Three broad classes of cloud – low, middle, and high – are defined by temperature as measured in AVHRR channel 4 and are further categorized by the underlying surface type. Not all surface/cloud level combinations occur in the study image, and those that do not are excluded from the analysis. Eight cloud classes are examined: low cloud over land, water, and ice; middle cloud over water and ice; and high cloud over land, water, and ice (abbreviations used: LCLL, LCLW, LCLI, MCLW, MCLI, HCLL, HCLW, and HCLI, respectively). The data are classified by two procedures: a neural network and a maximum likelihood classifier. The maximum likelihood procedure is supervised, initially using the same training areas as used to train the neural network.

The development of neural network architectures as a computational method builds on the fact that the brain is overall much more powerful in processing information than any serial computer. Neural networks (NN), connectionist systems, or parallel distributed processing (PDP) systems consist of networks of independent processors or units that are highly interconnected and process information through interaction between the individual processing nodes. Interaction between processors is determined by the network architecture: the number of layers, number of units in each layer, strengths of the connections (weights) between the units, unit activation functions, learning rules, and data representation. The advantage of PDP in the example presented here is the capability of such networks to learn by example. Through the use of learning rules, a neural network adjusts its connection weights to associate a set of input patterns with a set of output patterns and thereby "learns" the relationship between the input and output. In this study, multispectral data in training areas provide the input pattern; the desired cloud/surface class is the output pattern. Because processing in PDP systems is done through the interaction of many processing units, neural networks further display a feature known as "graceful degradation" where, given partially missing or noisy information, they are frequently capable of computing meaningful output. Of course, if classes are overlapping in feature space, and one or more of the features which best discriminate between classes are missing or otherwise corrupt, the network may produce in incorrect classification.

Forward-feed networks consist of a layer of input units, one or more layers of hidden units, and a layer of output units. The input units have direct connections to the units in the hidden layer, which in turn are connected to the output layer. Information processing flows from the input layer through the hidden layer to the output layer, and no feedback mechanism from the output to the input layer exists. There are also no direct connections between individual nodes within a layer. The relatively small number of connections, and therefore number of learnable connection weights, allows this type of network to train quickly while still being capable of solving complex problems through the construction of powerful generalizations.

A back-propagation network learns in the following way. An input and a training pattern – the "correct" output for a given input – are presented simultaneously to the network. Through

the connections, which are initially random, the network computes an output pattern, which is then compared to the training pattern. The error between the output pattern and the training pattern is used to adjust the weights between the output and hidden units to minimize the error between the training pattern and the output pattern. Because the output units are directly connected only to the hidden units, the training error is propagated backward through the network so that the weights between hidden and the input layer are also adjusted according to the learning rule (McClelland and Rumelhart, 1988). Training and input patterns are presented to the network many times while the network is adjusting its weights to minimize the error between all input and training patterns. The network will converge on a solution that maps the set of input vectors to the set of output vectors, if such a mapping function exists. During this training process, the units in the hidden layer construct generalizations or internal representations of the input patterns.

The network presented in this paper uses a forward-feed architecture with a layer of seven input units representing the AVHRR and SMMR channels, a layer of ten hidden units, and a layer of 12 output units representing the surface/cloud classes. The network is trained on patterns (training areas) for each desired class. After training, the network is presented with the complete data set (image) and computes a membership value for each pixel, represented by the activation of the output units.

Where spatial and spectral boundaries between phenomena are diffuse, hard classifiers which produce mutually exclusive classes seem particularly inappropriate. This issue is discussed further in Key *et al.* (1989) in relation to the fuzzy c-means algorithm. The neural networks approach addresses this problem of often indistinct spectral boundaries between classes by providing as output a numeric value as well as the class symbol for each pixel. This is a membership value for the pixel to each class, and is in the range [0,1], larger values implying greater strength of membership in a particular class. With the activation function used in this study (sigmoid), these values are approximately linear throughout the range.

In this example, two sets of training areas (referred to as TA1 and TA2) were selected in the typical supervised manner, with each training area manually delineated in the digital imagery. The training sets were chosen so that the effects of different within- and between-class variability could be tested, with TA1 representing a relatively small sample designed to study the ability of the neural network to address within-class variability not contained in the training statistics (e.g., the variability expected due to changes in ocean and ice temperature, ice concentration, and cloud height and thickness over space and time). TA1 included 1 percent of the 15,500 pixels (125 lines by 124 pixels) in the test data set. Additional training areas were included in TA2 to expand the variance of the training statistics sufficiently so that a significant portion of the test images would be classified using the maximum likelihood classifier. TA2 included about 9 percent of available pixels. Class means by spectral channel were nearly the same in TA1 and TA2 but, with the exception of the LCLI class, standard deviations were twice as large on average in TA2 (mean standard deviation in DN of 1.9 for TA1 versus 3.8 for TA2). As noted earlier, selection of training areas this large is not practical for climate applications requiring analysis of many images over large areas (thus, the impetus to test the neural network using TA1). However, TA2 was needed to address the trade-off between classification accuracy and human interaction using the supervised maximum likelihood approach for the types of data used here. The 1 July and 4 July images were manually interpreted using digital color composites of several AVHRR and SMMR channels. The manual interpretation thus acts as a hard classifier, with classes that consist of a "best-guess" estimate of class membership based on visual clues. Maximum likelihood (ML) classifications of the 1 July and 4 July images with seven data channels as input were carried out using TA1 and TA2 statistics. The ML procedure was run on a DEC MicroVax computer and required approximately 1 CPU minute for the computation of training area statistics and 2.5 minutes for image classification.

The neural network was trained using the individual pixel values in TA1 as input patterns. The network learned these sets of spectral patterns in approximately 23 minutes on an IBM PC/AT (12 MHz) class machine. The trained network then classified the entire study area in approximately 4 minutes. To address indistinct spectral boundaries, pixels were assigned to the class with the highest membership value. Pixels with no membership value greater than 0.4 (arbitrary) were tagged as unclassified. In an attempt to similarly relax the restrictions of the maximum likelihood classifier, cut-off limits for pixel position within the n-dimensional Gaussian class-membership curve were varied to a maximum of 99 percent, and different *a-priori* probabilities were tested.

RESULTS

Figure 1 illustrates many of the problems involved in mapping polar surfaces and clouds. With the exception of land/water boundaries, edges between classes are typically indistinct, as is the case at the sea ice margins where the transition from high concentration ice to open water is not distinct. For classes that include thin cloud, such as the regions of low cloud over ice and water, cloud opacity varies within the class. The sensors therefore record varying proportions of surface and cloud within a single class. Cloud height is also subjective; heights actually may fall anywhere within the low, medium, high designations.

A manual classification of the 1 July image with the classes described previously is shown in Figure 2, and will be used for comparison to the neural network and maximum likelihood results. Eighteen percent of the image was left unclassified where no dominant class could be determined. As noted above, within-

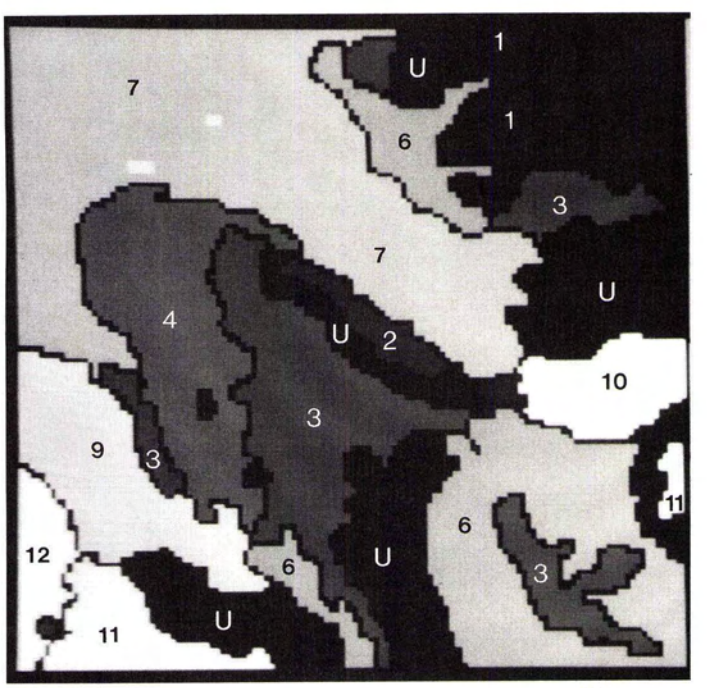

FIG. 2. Manual classification of the data shown in Figure 1. Classes are the same as in Figure 1 with the addition of U: unclassified.

class variance is large, particularly in classes LCLI and LCLW. Coefficients of variation are greatest in the 18 GHz SMMR data and for the AVHRR visible-wavelength channels within these classes, suggesting some confusion between ice and open water.

Table 1 shows the total percentages of the 1 July and 4 July images that were actually classified using training areas TA1 and TA2, ad the ML and NN classifiers. The neural network (NN) classification of the 1 July image is shown in Figure 3. Some important differences are apparent between the neural network output and the manual classification in the ICE and LCLI classes. The NN results underestimate the amount of low could over ice. The NN classification also puts a larger portion of the ice margin area into the WATER rather than ICE class. The contingency table (Table 2) comparing classifications of the 1 July image using NN and the manual interpretation confirm this, and also shows that NN tends to assign cloud/surface classes to surface classes, particularly in the case of ICE versus LCLI and WATER versus LCLW. Confusion also exists between cloud height classes. Of the pixels classified in both the NN and manual classifications (i.e., excluding unclassified pixels), overall agreement between classification schemes is 53 percent. The NN classification of the 4 July image shows similar patterns. As was the case for the 1 July data, nearly the entire image was classified. Differences between the NN results and manual classification for 4 July were greatest between cloud height classes and between low cloud over ice versus clear sky over ice.

TABLE 1. PERCENT OF IMAGES CLASSIFIED BY METHOD AND TRAINING SET.

Method	Training Set	Image	% Classified
ML	TA1	1 July	2
ML	TA1	4 July	3
ML	TA2	1 July	70
ML	TA2	4 July	53
NN	TA1	1 July	96
NN	TA1	4 July	93

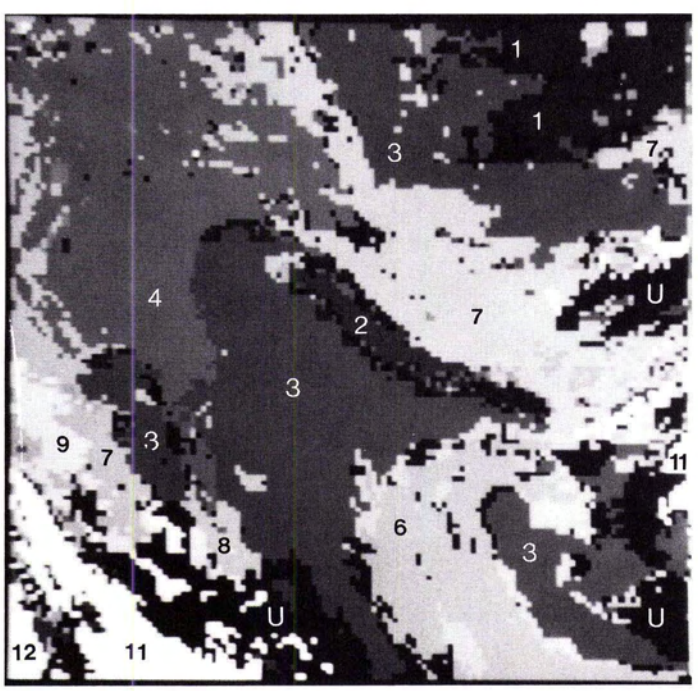

FIG. 3. Neural network classification of the study area. Classes are the same as in Figure 2. A pixel's class is the one in which it exhibited the largest membership value. A pixel is left unclassified if none of its membership values exceeds 0.4.

Maximum-likelihood classified images are shown in Figure 4 for 1 July using TA2 statistics. Because the ML classification using TA1 essentially included only those pixels within and adjacent to the TA1 training areas, these images are not illustrated. Comparing the manual interpretation in Figure 2 with the ML classification shows that the ML classification using the more comprehensive training areas of TA2 effectively captures the basic cloud and surface patterns. However, more than half of the manually-interpreted MCLW class is left unclassified by ML. Remaining unclassified pixels are divided among cloud classes and ice/water classes. A contingency table (Table 3) of the manual classification versus the ML classification using the TA2 training areas illustrates the problem of distinguishing between cloud height classes and between intermediate mixes of cloud and surface classes. This supervised ML classification achieved a high agreement of 85 percent, representing the large training set in TA2. Extension of these TA2 signatures to the 4 July data using ML illustrates the reduction in applicability of training signatures over time compared to the NN classifications, as shown by a general decrease in the percentage of the image that is classified. Given the variability of ice conditions and cloud thicknesses within a single image, it is not surprising that day-to-day variability is enough to reduce the representativeness of the training areas in terms of class mean and covariance. With the exception of class LCLI, standard deviations of training areas in TA1 are considerably less than is the case for TA2 and the manually interpreted classes. TA1 signatures thus include only a small portion of the variance in the desired classes, as indicated by the low percentage of the image actually classified using TA1 statistics.

DISCUSSION

The pattern of weights in the network provides insight into the way decisions are made by the network. For our example, interpretation of these weights sheds light on which spectral channels are important for classifying particular cloud/surface classes. Similarly, the network weights indicate, for each output class, the degree of information redundancy among channels in the input data; channels that are only weakly weighted add little additional information to the classification process. Figure 5 shows an example of the connections between the input channels, hidden layer neurons, and the output classes in the trained network used here. Due to the complexity of the connections between units, only the surface classes are shown in the figure. The identification of the exact role of hidden units is difficult, as they often represent generalizations of the input patterns. Strength of the connections shown varies from 0.6 to 0.9 (on a scale of -1.0 to $+1.0$). These connections are summarized in Tables 4 and 5. Table 4 shows with which input data channel each hidden node is associated in the trained network. Table 5 shows the association between hidden units and the output data classes. Following the connections through these two tables, therefore, indicates which input channels are linked to particular output classes. As shown in Figure 5 and Table 5, snow-free land has strong connections with hidden layer neurons, 2, 3, and 6, all of which represent thermal AVHRR channels and the SMMR channels (Table 4). We may therefore conclude that land is best distinguished from the other channels by its physical temperature and its emissivity in the microwave portion of the spectrum. Snow/ice cap is identified by its albedo and temperature, with no significant information gained from the microwave signature. The identification and discrimination of ice from the other classes requires albedo, temperature, and microwave emissivity characteristics. The connections show that

TABLE 2. CONTINGENCY TABLE SHOWING THE PERCENTAGE OF PIXELS IN THE IMAGE CLASSIFIED IN EACH OF THE TWELVE CLASSES BY THE MANUAL INTERPRETATION (HORIZONTAL) AND THE NEURAL NETWORK (VERTICAL) USING TRAINING SET TA1. PERCENTAGE = NUMBER OF PIXELS IN CLASS/TOTAL NUMBER OF CLASSIFIED PIXELS [12,251]. TOTAL AGREEMENT (SUM ALONG THE DIAGONAL) = 52.7%, BASED ON 1% OF THE IMAGE USED IN TRAINING AREAS.

NN Class	MANUAL CLASSIFICATION											
	LAND	SNOW	WATER	ICE	LCLL	LCLW	LCLI	MCLW	MCLI	HCLL	HCLW	HCLI
LAND	2.92	0.01	0.00	0.18	0.00	0.00	0.24	0.00	0.00	0.04	0.00	0.00
SNOW	0.00	1.01	0.00	0.00	0.00	0.00	0.18	0.00	0.04	0.15	0.00	0.00
WATER	0.58	0.58	14.41	2.47	0.00	4.04	0.58	1.40	0.27	0.05	0.15	0.03
ICE	0.01	0.06	0.00	6.17	0.00	0.00	13.46	0.00	0.08	0.01	0.00	0.08
LCLL	0.03	0.00	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.13	0.00	0.00
LCLW	0.00	0.03	0.00	0.00	0.00	4.90	0.03	3.01	0.10	0.38	0.01	0.00
LCLI	0.01	0.01	0.04	0.19	0.00	2.82	10.99	3.10	2.95	2.18	0.04	0.25
MCLW	0.00	0.00	0.00	0.00	0.00	3.69	0.00	6.29	0.86	0.13	0.38	0.00
MCLI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	2.09	0.25	0.01	0.04
HCLL	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
HCLW	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.19	0.00	2.40	0.01
HCLI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.17	1.45
TOTAL %	3.6	1.7	14.5	9.0	0.0	15.5	25.9	14.7	6.6	3.4	3.2	1.9
AGREEMENT (NN/TOTAL)	82.2%	59.4	99.7	68.5	0.0	31.7	42.4	42.7	31.8	0.0	75.9	78.0

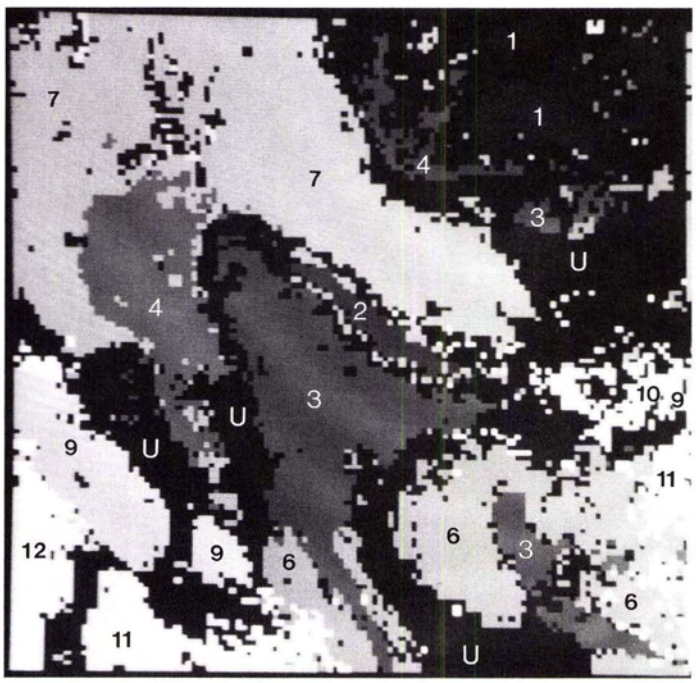

FIG. 4. Supervised maximum likelihood classification of the study area. Classes are the same as in Figure 2.

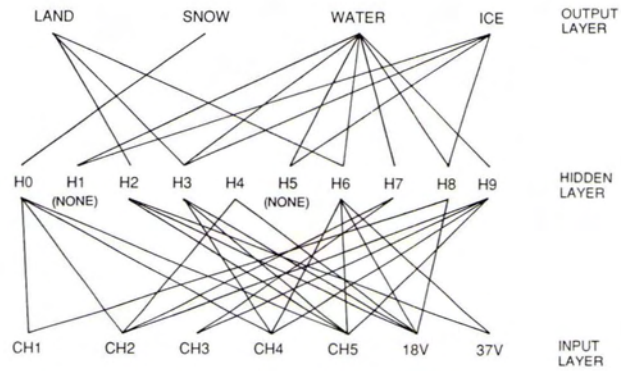

FIG. 5. Connections between the input channels, hidden layer neurons, and the output classes in the trained network. Output neurons represent the surface classes only. Strength of the connections shown varies from 0.6 to 0.9 (on a scale of −1.0 to +1.0). See also Tables 4 and 5.

cloud identification is a function of height, with thermal characteristics being more important for middle- and high-level clouds. The identification of low cloud depends on the underlying surface, where temperature is an adequate discriminator if the cloud is over land, albedo and temperature are used if over water, and temperature and the longer-wave reflected solar component (AVHRR channel 3) are needed if over ice. The hidden layer neurons with connections to AVHRR channel 3 (numbers 7 and 9) are also connected to either channel 2 or channels 4 and 5, indicating that both the reflected solar and thermal components of channel 3 likely play a part in the classification. The connections demonstrate the usefulness of AVHRR channel 3 for discriminating between cloud and snow or ice. Finally, note that hidden layer neurons 1 and 5 do not "listen" to any inputs, and therefore do not add any information to this network.

Although both the neural network and the ML classifier use the same training data, some fundamental differences exist in the way they are used. The neural network does not directly address the mean and covariance within a training area. Instead, each pixel within the training area is a separate pattern that directly influences the development of node weights. The multispectral characteristics of each pixel imprints itself to some degree on the network connections. During the development of unit weights as part of the network training phase, some aspects of training area means and covariances are included in the weight assignments. However, unlike the ML classifier, the neural network is not limited to assuming a statistical relationship between pixels within a class and is not restricted by assumptions of normality in the data. The fact that the multispectral data used here often violate these assumptions (Key, 1988) may contribute to the low percentage of the data classified using ML and the training statistics in TA1.

To test this hypothesis, a synthetic data set of AVHRR and SMMR data was developed that exhibited truly normal data dis-

TABLE 3. CONTINGENCY TABLE SHOWING THE PERCENTAGE OF PIXELS IN THE IMAGE CLASSIFIED IN EACH OF THE TWELVE CLASSES BY THE MANUAL INTERPRETATION (HORIZONTAL) AND THE SUPERVISED MAXIMUM LIKELIHOOD PROCEDURE (VERTICAL) USING TRAINING SET TA2. PERCENTAGE = NUMBER OF PIXELS IN CLASS/TOTAL NUMBER OF CLASSIFIED PIXELS [9,374]. TOTAL AGREEMENT (SUM ALONG THE DIAGONAL) = 84.8%, BASED ON 9% OF THE IMAGE USED IN TRAINING AREAS.

ML Class	MANUAL CLASSIFICATION											
	LAND	SNOW	WATER	ICE	LCLL	LCLW	LCLI	MCLW	MCLI	HCLL	HCLW	HCLI
LAND	4.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SNOW	0.00	1.37	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
WATER	0.00	0.00	13.64	0.05	0.00	1.46	0.00	0.08	0.00	0.00	0.00	0.00
ICE	0.00	0.00	0.00	7.56	0.00	0.00	0.23	0.00	0.02	0.00	0.00	0.00
LCLL	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LCLW	0.00	0.00	0.41	0.00	0.00	10.60	0.00	2.89	0.00	0.00	0.02	0.00
LCLI	0.02	0.12	0.00	1.09	0.00	0.17	29.38	0.07	0.10	0.12	0.00	0.00
MCLW	0.00	0.00	0.08	0.00	0.00	3.92	0.00	4.35	0.00	0.00	0.02	0.00
MCLI	0.00	0.00	0.00	0.00	0.00	0.02	0.33	0.36	5.79	0.45	0.07	0.05
HCLL	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.05	1.67	0.00	0.00
HCLW	0.00	0.00	0.03	0.00	0.00	0.64	0.00	1.54	0.00	0.00	3.29	0.00
HCLI	0.00	0.00	0.02	0.00	0.00	0.00	0.05	0.00	0.81	0.00	0.02	2.71
TOTAL %	4.5%	1.5	14.2	8.7	0.0	16.8	30.1	9.3	6.8	2.2	3.4	2.8
AGREEMENT (ML/TOTAL)	98.5%	91.9	96.1	86.9	0.0	63.1	97.6	48.8	85.5	74.6	96.2	98.2

TABLE 4. CONNECTIONS BETWEEN THE INPUT CHANNELS AND THE HIDDEN LAYER IN THE TRAINED NEURAL NETWORK.

Hidden Layer (Neuron number)	Input Channel Connections (AVHRR: 1,2,3,4,5; SMMR: 18,37)
0	1, 2, 4, 5
1	NONE
2	5, 18, 37
3	4, 5, 18
4	2, 18
5	NONE
6	4, 5, 18, 37
7	2, 3
8	1, 18
9	2, 3, 4, 5

TABLE 5. CONNECTIONS BETWEEN THE OUTPUT CLASS AND THE HIDDEN LAYER IN THE TRAINED NEURAL NETWORK.

Output Layer (Class)	Hidden Layer (Neuron number)
LAND	2, 3, 6
SNOW	0
WATER	1, 3, 5, 6, 7, 8, 9
ICE	1, 3, 5, 8
LCLL	6
LCLW	0, 9
LCLI	9
MCLW	1, 4, 5, 7, 9
MCLI	0, 4, 5, 9
HCLL	2, 4, 7
HCLW	1, 4, 5, 7, 8
HCLI	0, 3, 4, 8

tributions. Rectangular cloud and surface objects of varying sizes and locations were generated whose dimensions were randomly chosen within a restricted image. Object regions were then filled with normally distributed data for each channel based on pre-specified means and standard deviations (Gaussian random number generator) characteristic of the polar clouds and surfaces. With this data set there will be only one statistical class for each physical class (e.g., land, low cloud over water, etc.). For this reason, and because the data are normally distributed with a known variance, the probability of selecting a training area representative of the population is higher than with the actual data. Therefore, even small training areas should provide enough information about each class to allow a larger proportion of the area to be classified. This was in fact the case, where training areas extracted from less than 1 percent of the synthetic image allowed approximately 70 percent of the image to be correctly classified by the ML procedure. This test suggests that deviations from a normal distribution likely contribute to the low percentages of classification using the ML classifier.

The ability of a neural network to compute similarity measures through a comparison of patterns contributes to its ability to classify large portions of data on two separate images. Thus, although a relatively small portion of the variability of clouds and surfaces were captured in the training areas (particularly TA1), the neural network was still able to reliably choose the most appropriate output class. This property provides a means to address the problem of signature extension over time and space, since a properly trained network can make class assignments—albeit with reduced confidence—in the face of atmospheric effects or slight changes in spectral properties without requiring *a-priori* knowledge of within-class variance or probabilities of class membership. In fact, if one has a particular reason to use a statistical classifier, the strength-of-membership values calculated by a neural network could be fed back into the statistical classifier as apriori probabilities using, for example, the probability image feature in the ERDAS™ MAXCLAS function.

The classification example presented has utilized numeric data—albedos and brightness temperatures—as input. Output is on the nominal level for both the neural network and the ML procedure, although the neural network also provides a type of membership value. In some cases, input such as category identifiers rather than measurements may be useful where pixels are assigned a class symbol and optionally an associated fuzziness (e.g., the probability that the pixel belongs to the class). A second neural network was developed that uses both nominal and categorical input. For example, in the study area the locations of land and permanent ice cap are known, and the location and concentration of sea ice can be determined from the SMMR data. Consider also the case where only three broad categories of sea ice concentration are of interest: low (15 to 40 percent), medium (41 to 70 percent), and high (71 to 100 per-

cent). Other variables are also possible; for example, time of year, geographic location, spatial context, texture, stage of plant growth, and the *a priori* probability of occurrence of each surface or cloud type. For simplicity, however, the example is limited to symbols representing land/not land, ice cap/not ice cap, and low, medium, and high sea ice concentration variables.

This network was trained with input units corresponding to both spectral and categorical variables. Because some of the AVHRR and SMMR channels are highly correlated—as evidenced from principal components analysis and an examination of the previously described neural network, only AVHRR channels 1, 3, and 4, and SMMR channel 18 GHz vertical were used in the training. Categorical input variables represent land, ice cap, and sea ice concentration. Nine hidden layer units were specified with output classes as before. As expected, the resulting classification (not shown) is similar to that using only spectral information but the proportion of correctly identified surface pixels increased slightly, whereas the proportion of cloudy pixels remained essentially the same. In addition, the certainty with which surface pixels were classified as measured by the output membership values increased significantly; with some coastal pixels the increase was as much as 0.4.

CONCLUSIONS

Four surface and eight cloud categories in merged AVHRR and SMMR data for the Arctic summer were identified through a neural network approach and a traditional maximum likelihood procedure. Both the numeric and the numeric/symbolic neural networks extracted correct information from the multispectral images. The differences between the neural network and the supervised maximum likelihood classifications were primarily due to the greater flexibility of the neural network to classify indistinct classes, e.g., classes containing pixels with spectral values that differ significantly from those in the training areas, while ignoring assumptions of statistical normality. The two classification approaches illustrate the tradeoffs between human interaction in the selection of training areas and classification accuracy and flexibility. Flexibility similar to that shown by the neural network might be achieved using a maximum likelihood routine by manipulating class membership probabilities and/or by adjusting probability thresholds to relax the membership requirements for individual classes. Such steps may require an *a priori* knowledge of probabilities or may increase classification error. In any case, such tuning was not effective in the example described here.

The neural network approach to classification is generally less rigid than the traditional maximum likelihood procedure in that (1) there are no assumptions of distributions of variables and relationships between them, (2) the network is easily trained to learn the relationships between input and output, and (3) the classification produces both a categorical value and a type of membership value for each pixel. It is recognized that there is some loss of information and interpretability with the departure from statistical theory. Additionally, computation time required for training the network is not trivial when compared to the training of the ML classifier (i.e., computation of mean vectors and the covariance matrix), although future hardware architectures should alleviate this problem. Of course, training time as a proportion of the total classification time decreases with the amount of data processed, so that if classes do not change and large images are being classified, overall processing time should be similar for both methods.

The ability to interpret weights within the trained network provides a potentially powerful tool for understanding the role of inputs and the geophysical processes they represent in the making of decisions. Through an examination of the connection strengths between input, hidden, and output units, it is possible to identify which inputs influence the classification most, and which are redundant. These relationships are not always clear, and care must be taken in extending their interpretation to physical processes. It was also shown that ancillary information, even on a simplistic level, can improve classification accuracy and can be easily included in a network. Although the example provided indicated that maximum likelihood results could be made to agree more closely with the manual interpretation, this was achieved only after training areas were expanded to include 9 percent of the test image. Such a degree of training is impractical for remote sensing climate studies because of the volume of imagery that must be processed. We emphasize that the data and applications of interest for remote sensing of polar climate are not typical of applications such as land-cover mapping, which may be limited to a single image covering relatively small areas with small within-class variance. This study does not show whether a neural network offers any advantages for the latter type of analyses. The merits and drawbacks of a neural network approach relative to others must therefore be considered based on the particular problem at hand.

ACKNOWLEDGMENTS

This work was supported by NASA grants NAG-5-898 and 3455-OP-610. Thanks are due to W. Rossow and E. Raschke for providing AVHRR GAC data.

REFERENCES

Andress, K., and A. Kak, 1988. Evidence accumulation and flow of control. *AI Magazine*, 9(2), 75–94.

Barr, A., and E. A. Feigenbaum, 1981, 1982. *Handbook of Artificial Intelligence*. 3 vols. William Kaufmann, Inc., Los Altos, California.

Borchardt, G. C., 1986. *STAR:* a computer language for hybrid AI applications. *Coupling Symbolic and Numerical Computing in Expert Systems* (J. S. Kowalik, ed.), North-Holland, Amsterdam, The Netherlands.

Campbell, W. J., and L. H. Roelofs, 1984. Artificial intelligence applications for the remote sensing and earth science community. *Proceedings, Ninth Pecora Symposium on Spatial Information Technologies for Remote Sensing Today and Tomorrow*, Sioux Falls, South Dakota.

Chandrasekaran, B., A. Goel, and D. Allemang, 1988. Connectionism and information processing abstractions. *AI Magazine*, 9(4), 24–34.

Ebert, E., 1988. *Classification and Analysis of Surface and Clouds at High Latitudes from AVHRR Multispectral Satellite Data. Scientific Report #8*, Dept. of Meteorology, Univ. of Wisconsin-Madison.

———, 1987. A pattern recognition technique for distinguishing surface and cloud types in the polar regions, *J. Clim. Appl. Meteor.*, 26, 1412–1427.

Estes, J. E., C. Sailer, and L. R. Tinney, 1986. Applications of artificial intelligence techniques to remote sensing. *Professional Geographer*, 38(2), 133–141.

Holmes, Q. A., D. R. Nuesch, and R. A. Schuchman, 1984. Textural analysis and real-time classification of sea-ice types using digital SAR data. *IEEE Trans. Geosci. Rem. Sens.*, GE-22(2), 133–120.

Jackson, M. J., and D. C. Mason, 1986. The development of integrated geo-information systems. *Int. J. Rem. Sens.*, 7(6), 723–740.

Key, J. R., 1988. *Cloud Analysis in the Arctic from Combined AVHRR and SMMR Data*. Ph.D. dissertation, Department of Geography, University of Colorado, Boulder, 180 p.

Key, J. R., J. A. Maslanik, and R. G. Barry, 1989. Cloud classification using a fuzzy sets algorithm: a polar example, *Int. J. Rem. Sens.* (in press).

Kitzmiller, C. T., and J. A. Kowalik, 1987. Coupling symbolic and numeric computing in KB systems. *AI Magazine*, 8(2), 85–90.

Kowalik, J. S. (ed.), 1986. *Coupling Symbolic and Numerical Computing in Expert Systems*. North-Holland, Amsterdam, The Netherlands.

Lauritsen, L., G. G. Nelson, and R. W. Port, 1979. *Data Extraction and Calibration of TIROS-N/NOAA A-G Radiometer*, NOAA Tech. Memor., *NESS* 107, Natl. Oceanic and Atmos. Admin., Boulder.

Maslanik, J. A., J. R. Key, and R. G. Barry, 1989. Merging AVHRR and

SMMR data for remote sensing of ice and cloud in polar regions. *Int. J. Rem. Sens.* (in press).

Matsuyama, T., 1987. Knowledge-based aerial image understanding systems and expert systems for image processing. *IEEE Trans. Geosci. Rem. Sens.*, GE-25 (3), 305–316.

McClelland, J. L., and D. E. Rumelhart, 1988. *Explorations in Parallel Distributed Processing*. MIT Press, Cambridge, Mass. 344 p.

McKeown, D. M., Jr., 1987. The role of artificial intelligence in the integration of remotely sensed data with geographic information systems. *IEEE Trans. Geosci. Rem. Sens.*, GE-25 (3), 330–348.

Minsky, M., and S. Papert, 1969. *Perceptrons*. MIT Press, Cambridge, Mass.

Moravec, H., 1988. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, 9(2), 61–74.

Murthy, H. A., and S. Haykin, 1987. Bayesian classification of surface-based ice-radar images. *IEEE J. Ocean. Eng.*, OE-12 (3), 493–502.

Nandhakumar, N., and J. K. Aggarwal, 1985. The artificial intelligence approach to pattern recognition-a perspective and an overview. *Pattern Recognition*, 18 (6), 383–389.

Nicolin, B., and Gabler, 1987. A knowledge-based system for the analysis of aerial images. *IEEE Trans. Geosci. Rem. Sens.*, GE-25 (3), 317–328.

NOAA, 1984. *NOAA polar orbiter data user's guide*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, NESDIS, February.

Reddy, R., 1988. Foundations and grand challenges of artificial intelligence. *AI Magazine*, 9 (4), 9–21.

Ripple, W. J., and V. S. Ulshoefer, 1987. Expert systems and spatial data models for efficient geographic data handling. *Photogrammetric Engineering and Remote Sensing*, 53 (10), 1431–1433.

Ritter, N. D., T. L. Logan, and N. A. Bryant, 1988. Integration of neural network technologies with geographic information systems. *GIS Symposium: Integrating Technology and Geoscience Applications*, September, Denver, Colorado, pp. 102–103.

Robinson, V. B., and A. U. Frank, 1987. Expert systems for geographic information systems. *Photogrammetric Engineering and Remote Sensing*, 53 (10), 1435–1441.

Rumelhart, D. E., J.L. McClelland, and the PDP Research Group, 1986. *Parallel Distributed Processing*. MIT Press, Cambridge, Mass. 547 p.

Schwalb, A., 1984. *The TIROS-N/NOAA A-G Satellite Series*, NOAA Tech. Mem., *NESS 95*.

Smith, T. R., 1984. Artificial intelligence and its applicability to geographical problem solving. *Professional Geographer*, 36 (2), 147–158.

Smith, T., D. Peuquet, S. Menon, and P. Agarwal, 1987. KBGIS II: a knowledge-based geographical information system. *Int. J. Geog. Sys.*, 1 (2), 149–172.

Usery, E. L., and P. Altheide, 1988. Knowledge-based GIS techniques applied to geological engineering. *Photogrammetric Engineering and Remote Sensing*, 54 (11), 1623–1628.

WMO, 1987. *Report of the ISCCP Workshop on Cloud Algorithms in the Polar Regons*, World Climate Research Programme, WCP-131, WMO/TD-No. 170, Tokyo, Japan, 19-21 August 1986.