A Neural Network Approach to Cloud Classification

JONATHAN LEE, RONALD C. WEGER, SAILES K. SENGUPTA, AND RONALD M. WELCH, MEMBER, IEEE

Abstract—**Recent cloud retrieval validation studies suggest that there are major discrepancies between the various algorithms. Other studies have demonstrated that the use of texture-based pattern recognition features can significantly improve cloud identification accuracy. However, the capabilities and accuracies which can be attained with spatial information remain poorly understood and undocumented, and the choice of an optimal feature set is unknown.**

The results from this study demonstrate that, using high spatial resolution data, very high cloud classification accuracies can be obtained. A texture-based neural network classifier using only single-channel visible LANDSAT MSS imagery achieves an overall cloud identification accuracy of 93%. It is remarkable that cirrus can be distinguished from boundary layer cloudiness with an accuracy of 96%, without the use of an infrared channel. Stratocumulus is retrieved with an accuracy of 92%, cumulus at 90%. The use of the neural network does not improve cirrus classification accuracy. Rather, its main effect is in the improved separation between stratocumulus and cumulus cloudiness.

While most cloud classification algorithms rely on linear, parametric schemes, the present study is based on a nonlinear, nonparametric four-layer neural network approach. Intercomparisons are made to a three-layer neural network architecture, the nonparametric *K*-nearest neighhor approach, and the linear stepwise discriminant analysis procedure. A significant finding is that significantly higher accuracies are attained with the nonparametric approaches using only 20% of the database as training data, compared to 67% of the database in the linear approach.

I. Introduction

There is a rich history of cloud classification approaches, as outlined by Hughes [1], Goodman and Henderson-Sellers [2], and Rossow [3]. A crucial issue is accurate cloud identification. While the number of approaches to cloud classification continually increases, there have been few objective attempts towards cloud validation. However, Stowe *et al.* [4] report that for an identical global data set, the ISCCP and NCLE (Nimbus-7) algorithms produce a tremendous 17% discrepancy in global cloud cover. To place this in perspective, Hansen *et al.* [5] estimate that a 4% increase in low cloudiness would be sufficient to totally mask the radiative effect of a doubling of CO_2 in the atmosphere. Even more alarming are the detailed regional studies of Parker and Wielicki [6] in which seven different cloud retrieval algorithms are intercompared as a function of spatial resolution. They show that the ISCCP algorithm overestimates stratocumulus cloudiness by about 17%, while the NCLE algorithm underestimates stratocumulus by about 20%; likewise, for cirrus cloudiness, ISCCP overestimates by 7%, while NCLE underestimates by 38%. Clearly, accurate cloud identification has not yet been demonstrated.

The present study takes a very different approach, demonstrating that very high cloud classification accuracies can be achieved using textural measures derived from a single *visible* channel of high spatial resolution imagery. It is remarkable that high cirrus clouds have textural features which distinguish them from boundary layer cloudiness, without the use of infrared channel information. This investigation is focused upon determining the maximum cloud classification accuracy which can be attained from single visible channel imagery. The methodology is to apply the neural network approach.

Neural networks have been touted as having excellent potential for improving classification accuracy in geophysical data. However, there have been few studies which have demonstrated this potential using real data sets. The appeal of neural networks as pattern recognition systems is based upon several considerations. First, neural networks appear to perform as well or better than other techniques, and require no assumptions about the explicit parametric nature of distributions of the pattern data to be classified (e.g., normal distributions). In this regard, they are similar to *K*-nearest neighbor algorithms. However, neural networks, once trained, are computationally more efficient. Finally, a comparison of neural networks to *K*-nearest neighbor and discriminant analysis methods indicates that neural networks can achieve a parity of performance using a much smaller set of training data.

Section II outlines how neural networks have been used in other disciplines. Section III describes the neural network approach to pattern recognition, and Section IV provides the neural network methodology applied to texture-based cloud classification. Section V describes the *K*-nearest neighbor approach. Section VI provides the cloud classification intercomparison results for the neural network, *K*-nearest neighbor, and stepwise discriminant analysis approaches, and Section VII concludes the paper.

II. Background

The neural network approach is attractive because it has the capability to learn patterns whose complexity makes them difficult to define using more formal approaches.

Manuscript received June 6, 1989; revised February 12, 1990. This work was supported in part by the National Science Foundation under Grant ATM-8816052 and in part by the National Aeronautics and Space Administration under Grant NAG-1-542.

J. Lee, R. C. Weger, and S. K. Sengupta are with the Department of Computer Science, South Dakota School of Mines and Technology, Rapid City, SD 57701.

R. M. Welch is with the Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, SD 57701, on leave at the Naval Oceanographic and Atmospheric Research Laboratory, Monterey, CA 93943.

IEEE Log Number 9037443.

0196-2892/90/0900-0846$01.00 © 1990 IEEE

Also, it has the ability to integrate information from multiple sources and to incorporate new features without degrading prior learning. Neural networks can generate solutions to complex problems which are beyond formal description. This is accomplished by developing knowledge about the problem by training on the pattern category in the form of feature values shared by members of the category, together with allowed tolerance for those features.

For example, the Neocognitron of Fukushima [7] is a massively parallel multilevel feedforward neural network which classifies two-dimensional patterns. It has been extended by workers at the MIT Lincoln Laboratory to successfully classify images by extracting features and retaining only those whose response exceeded an appropriate threshold.

Gorman and Sejnowski [8] used a three-layer neural network architecture with a backpropagation algorithm [9] to classify sonar returns from a mine and a "mine-shaped" rock positioned on a sandy ocean floor. Among other uses of neural network decision systems are risk analyses [10] in insurance underwriting, and in the development of "recognition automata" by Reeke and Edelman [11]. However, there are few studies using neural networks with geophysical data.

III. Pattern Recognition Using Neural Networks

A neural network consists of objects called nodes, an abstraction of the biological concept of a neuron, and weighted paths connecting these nodes. Each node has an activity represented by a real number. This activity value is typically computed as some nonlinear bounded monotone increasing function of a weighted sum of the activities of other nodes that are directly connected to the given node.

The neural network has three or more processing layers: an input layer, one or more hidden processing layers, and an output layer. The activity at an input node represents the value of some input signal. If the activity of node K is denoted by V_K, and a weight on a path from node L to node K is denoted by W_{KL}, then one possible formula for the activity at node K is given by

$$V_K = f\left(\sum_L W_{KL} V_L\right) \tag{1}$$

where f is an appropriately chosen (possibly) nonlinear function.

For the output nodes, a neural network may be viewed as a nonlinear vector-valued function:

$$\vec{O} = \vec{F}(\vec{I}) \tag{2}$$

where $\vec{O}$ is a vector with one component for each activity of an output node, and $\vec{I}$ is a vector with one component for the activity for each input node. From this viewpoint, neural networks are mechanisms for performing computation.

Neural networks may be classified according to the topology of the paths connecting the nodes and by the algorithm employed in determining the appropriate value for the path weights. In the architecture employed in pattern recognition in this paper, the paths form a loop-free directed graph. Such a structure is referred to as a "feedforward" network. The nodes are then grouped into layers, a layer being all nodes separated from an input node by a fixed number of arcs. The three-layer and four-layer feedforward neural network architectures used in the present study are shown in Fig. 1.

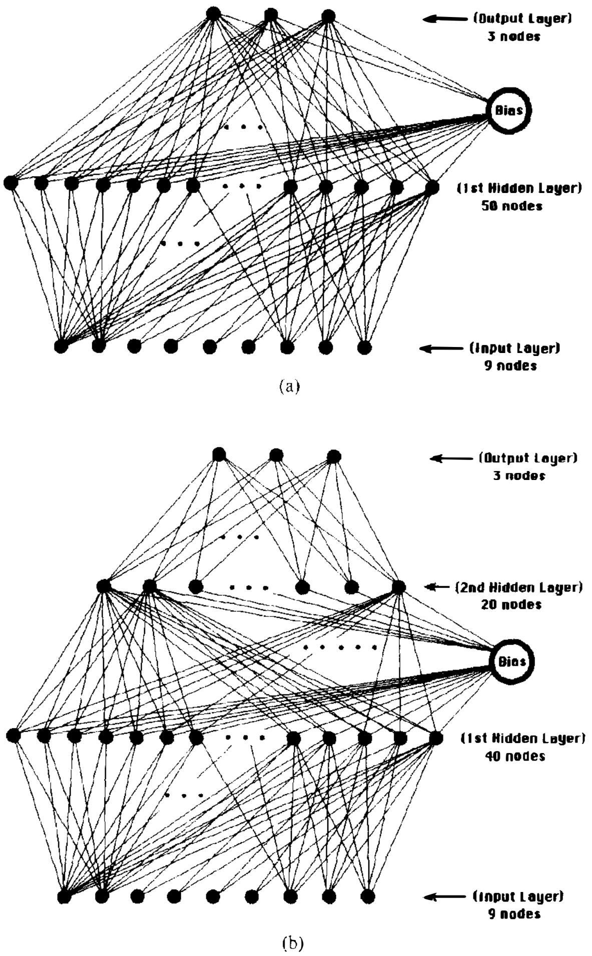


Fig. 1 Three-layer and four-layer feedforward neural network architectures.

The determination of appropriate weights W_{KL} is referred to as learning or programming. Learning algorithms may be classified as supervised or unsupervised. In the supervised learning mode, for each possible input vector $\vec{I}$, an associated output vector $\vec{O}$ is specified. The function of the learning algorithm is to choose the value of the weights so that $\vec{F}(\vec{I})$ is a good approximation to $\vec{O}$.

This mode of learning is appropriate in pattern recognition when 1) objects to be recognized are encoded as input vectors $\vec{I}_1, \vec{I}_2, \cdots, \vec{I}_N$, and 2) it is desirable to mechanically assign each object to exactly one of a predetermined set of pattern classes. If the pattern classes are encoded as output vectors $\vec{O}_1, \vec{O}_2, \cdots \vec{O}_M$, then the adequacy of the network as a pattern recognizer may be quantified by

$$E = \sum_{L=1}^{N} \left\| \vec{O}_L^* - \vec{F}(\vec{I}_L) \right\|^2 \tag{3}$$

where $\vec{O}_L^*$ encodes the pattern class to which object $\vec{I}_L$ belongs and $\| \cdot \|$ denotes the length of a vector. E is a measure of global error representing the sum of the squared classifier error at each input.

Rumelhart *et al.* [9] devised an algorithm referred to as "backpropagation" to iteratively determine weights W_{KL} that locally minimize E. The algorithm is a special case of a gradient search in which the weights are initialized as small random numbers, and are then repeatedly updated at the nth iteration according to the rule

$$\Delta \vec{W} = -\eta \vec{\nabla} E, \qquad \vec{W}_{n+1} = \vec{W}_n + \Delta \vec{W} \tag{4}$$

where $\vec{W}$ is a vector composed of the weights and $\vec{\nabla} E$ denotes the gradient of the error E as a function of the weights. η is a small positive parameter called the learning rate. As $\vec{\nabla} E$ is a vector in weight space pointing in the direction along which E increases most rapidly, $-\eta \vec{\nabla} E$ points in the direction of the steepest descent of E.

Unfortunately, if the gradient search process encounters a local minimum of E, then $\vec{\nabla} E = \vec{O}$ and $\vec{W}_{n+1} = \vec{W}_n$. A number of techniques may be employed to "unstick" the search process. A so-called "momentum" term may be added to give

$$\Delta \vec{W} = -\eta \vec{\nabla} E + M(\vec{W}_n - \vec{W}_{n-1}) \tag{5}$$

where M is a parameter which measures the influence of the previous search direction. In the case of gradient search applied to neural networks, increasing the numbers of layers appears to sometimes alleviate the problem.

IV. A Cloud Identification Network

In this paper, the problem of identifying digital satellite images of clouds is considered using a feedforward network trained by backpropagation.

A. Data Description

Landsat Multispectral Scanner (MSS) digital data with a spatial resolution of 57 m and a gray level range of 0–127 are used in this study. Each scene is composed of 3246 × 2983 pixels covering a land area 185 km wide × 170 km long. A total of 37 Landsat MSS scenes are used in this investigation—15 stratocumulus, 10 cumulus, and 12 cirrus.

It should be explicitly stated that the methodology of supervised classification is well established and is based on the assumptions: 1) that the existing body of knowledge can label each training sample into one of several known classes, and 2) that the test data come from one of these known classes. Clearly, the classification accuracy for any problem, regardless of methodology, it at best only as good as the proper assignment of labels. In the present case, a large number of scientists have examined the data set. The stratocumulus scenes range from nearly solid decks to "breakup regions" which somewhat resemble cumulus. The cumulus scenes range from small fair-weather cumulus over the ocean to large mesoscale-sized features over land. The cirrus scenes include cirrus, cirrostratus, cirrocumulus, and even contrails. These cloud types represent the most common and radiatively most important cloud types. Due to their dominating influence on the Earth's radiation balance, the First ISCCP Regional Experiment (FIRE) is concentrating exclusively upon cirrus and stratocumulus cloudiness. Three representative scenes from each of these three cloud classes are shown in Fig. 2. Each of the 37 total cloud scenes is subdivided into 20 subregions. Each subregion of 512 × 512 pixels (29 × 29 km) represents a sample from that scene. The area covered by the 37 scenes is equivalent to 0.3% of the Earth's surface. It is not at all unusual to find sampling rates of this size in the spatial statistics literature.

B. Definition of Textural Features

Texture is often interpreted in the literature as a set of statistical measures of the spatial distribution of gray levels in an image. The gray level co-occurrences matrix (GLCM) method assumes that the texture information in an image is contained in the overall or "average" spatial relationships that gray levels have with one another [12]. Our previous investigations [13], [14] showed that GLCM-based textural features are capable of discriminating cloud types. Cloud classification accuracies of 83–88% are achieved using this approach [15]. These studies demonstrate that there is high information content in the cloud spatial patterns.

Perhaps the principal drawback of many texture methods and the GLCM approach, in particular, is the requirement of large memory and computation time. A comparison of accuracy, computation time, and storage requirements for cloud classification was made between the GLCM, sum and difference histogram (SADH) approach of Unser [16] and the gray level difference vector (GLDV) approach of Weszka *et al.* [17]. It was found that the SADH and GLDV methods have cloud classification accuracies similar to those obtained using GLCM, but with significantly reduced requirements for both computation time and memory [18]. The GLDV approach is used in the present study.

C. Gray Level Difference Vector Features

GLDV is based on the absolute differences between pairs of gray levels I and J found at a distance d apart at an angle ϕ with a fixed direction. The difference vector probability density function $P(m)_{d,\phi}$ is defined for $m = |I - J|$, and is estimated by dividing the gray level frequencies of occurrence by the total frequencies. From this

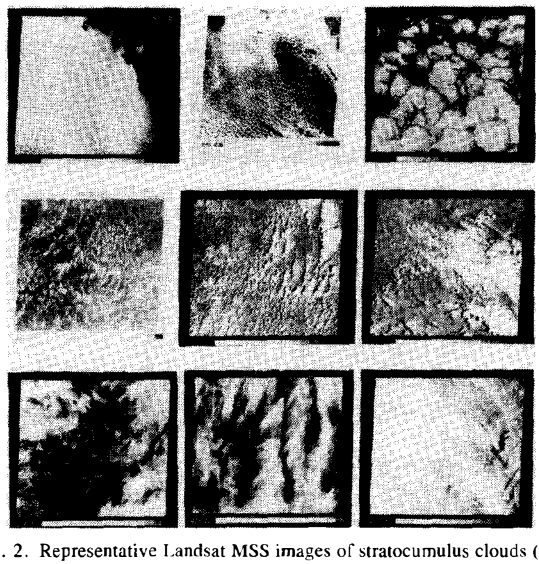


Fig. 2. Representative Landsat MSS images of stratocumulus clouds (first row), cumulus clouds (second row), and cirrus clouds (third row) used in this study. Each region is 185 km horizontally and 170 km vertically

density function, the individual textural measures are computed, as described below.

Mean: $\mu_{d,\phi} = \sum_m mP(m)_{d,\phi}$ (6)

Standard Deviation: $\sigma_{d,\phi}$

$$= \left[\sum_m (m - \mu_{d,\phi})^2 P(m)_{d,\phi}\right]^{1/2} \qquad (7)$$

Contrast: $\mathrm{CON}_{d,\phi} = \sum_m m^2 P(m)_{d,\phi}$ (8)

Angular Second Moment: $\mathrm{ASM}_{d,\phi}$

$$= \sum_m \left[P(m)_{d,\phi}\right]^2 \qquad (9)$$

Entropy: $\mathrm{ENT}_{d,\phi} = -\sum_m P(m)_{d,\phi} \log P(m)_{d,\phi}$ (10)

Local Homogeneity: $\mathrm{HOM}_{d,\phi}$

$$= \sum_m P(m)_{d,\phi}/[1 + m^2] \qquad (11)$$

Cluster Shade: $\mathrm{SHAD}_{d,\phi}$

$$= \left[\sum_m (m - \mu_{d,\phi})^3 P(m)_{d,\phi}\right] \Big/ \sigma^3_{d,\phi} \qquad (12)$$

Cluster Prominence: $\mathrm{PROM}_{d,\phi}$

$$= \left[\sum_m (m - \mu_{d,\phi})^4 P(m)_{d,\phi}\right] \Big/ \sigma^4_{d,\phi} - 3 \qquad (13)$$

Number of Cloud Pairs:

NOP is the number of cloud pixel pairs separated by distance d and orientation ϕ.

Plots of representative cloud texture measures as a function of pixel separation distance d and angle ϕ are shown by Welch *et al.* [15], [14] and Kuo *et al.* [13].

D. Procedure

A step-by-step outline of the process of using a neural network as a pattern classifier is now presented.

1) Encoding the Input and Output Patterns: The components of the input vectors should be normalized to lie in some fixed range. In the present study, each component of $\vec{I}$ was normalized to the interval [0, 1]. This prevents noise in a large component from swamping information in a smaller component.

Normalization is used to ensure faster convergence to a global optimum. It should be noted here that the results from the classifier consist of two parts:

a) the output vector determining the classification type of the test data, and

b) the relative significance of the feature components in determining the classification types.

In a), the results remain unaffected by the normalization provided convergence to the global minimum occurs in the unnormalized case. In b), the ranks of the absolute values and the signs of the terminal weights that determine the relative importance of the components in classification remain the same in the normalized and unnormalized cases.

The output vectors should be encoded so that each component lies within the range of the nonlinear function $f(\,)$. It is desirable to spread the output vectors so as to maximize the angle between individual vectors subject to the above constraint. In the present case, the images represented three distinct categories of clouds: stratocumulus, cumulus, and cirrus. The output vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

were chosen to encode these three categories.

2) Select a Network Topology and a Learning Algorithm: For supervised learning, feedforward networks using the backpropagation algorithm have been generally successful and enjoy wide popularity. Alternative learning strategies such as Hecht-Nielsen's [19] counterpropagation or approaches such as those used by Hopfield [20], [21] to learning on a network with an unrestricted topology have also been successfully employed.

Counterpropagation has been employed on two- or, at most, three-layer networks, and the thermodynamic models have a longer learning time. Thus, backpropagation on a multilayer feedforward network was selected for the present study. As noted above, increasing the number of layers beyond three appears to lessen problems with local minima of E. In this study, both three- and four-layer networks were investigated, with the best results obtained consistently with four layers. The number of nodes in the input and output layers are, of course, constrained by the number of variables in the input data and the number of output classes, respectively. Layers between the input and output layers are called "hidden" layers. There appears to be no simple way of determining the number of nodes required for these layers to function properly. The size is problem dependent. That is, in a given problem, there is a minimum number below which the learning algorithm is unable to minimize E adequately [28].

Although Kolomogorov has shown that three-layer networks are sufficient for representing continuous mappings of the form $\vec{O} = \vec{F}(\vec{I})$, finding such a representation in practice may prove difficult [22]. Note that the configuration of the network is problem dependent in the sense that a neural network that recognizes a handwritten signature has a different hidden layer configuration from one which classifies clouds. For a cloud classification problem, a given hidden layer with a fixed configuration is adequate. The exact same configuration is used in all cloud classifications in this paper, and is of a generic nature applicable to all cloud classification problems with the same number of inputs and outputs.

3) Training the Network: In neural networks, the learning phase is initiated by selecting a representative set of pattern examples and converting these examples into the encoded vector pairs (I_L, O_L). Beginning with small randomized weights, the weights are iteratively adjusted to minimize the global error E.

At this stage, there are two further issues.

1) How many training pairs are necessary?

2) In what manner should the training samples be presented to the network?

The answer to both questions depends upon the ease of class separation of the input vectors, the amount of noise present in the input signals, and the extent to which the set of input vectors belonging to one class of patterns overlaps another pattern category. In the presence of significant noise, it is undesirable to drive global error E too small. Therefore one should not iterate too often over the same training set.

When the sets of input vectors belonging to distinct classes overlap, a larger fraction of the available data will be required for satisfactory network performance. This results from the fact that the feedforward network distinguishes pattern classes by separating the various classes by nonlinear hypersurfaces.

V. The K-Nearest Neighbor (KNN) Classifier

Many classification methods assume that the form of class-conditional densities is known. The popular maximum likelihood estimation approach assumes multivariate normality. However, it is widely recognized that this assumption is suspect for textural features based on cloud pixel brightness.

The KNN procedure is a nonparametric classification procedure. This rule classifies a new feature vector $\vec{x}$ by assigning it the label most frequently represented among the K-nearest of all training samples [23]. The decision is made by determining the majority class represented in the set of K-nearest neighbors of a pattern by examining the labels of each of the K neighbors. Randomization is used for breaking ties. In practice, one chooses $K = c\sqrt{n}$ where c is an appropriate constant and n is the size of the training set. In the present study, $c = 1$ is used.

KNN classification is generally used for estimating the optimal (Bayes) error bounds [24]. However, we have used it in the present study as a nonparametric classifier for the purpose of comparison to the neural network classifier.

VI. Results

In a traditional classifier,

1) inputs and outputs are passed serially,

2) internal computations are performed sequentially, and

3) parameters are estimated from training patterns and then held at those estimated values as constants.

In contrast, in a neural net classifier,

1) both inputs and outputs can be performed in parallel, and

2) internal parameters (weights) are adjusted using the output and the appropriate training labels associated with them.

Using the expressions listed in Section IV, textural features for subregions and the corresponding databases consisting of those textural features are constructed. Results for the discriminant analysis, neural network, and K-nearest neighbor approaches are given below.

A. Discriminant Analysis

Two-thirds of the data set from the textural feature database are used to generate the training data for the classifier. The remainder are used as test data to determine classifier accuracy. Stepwise discriminant analysis [25], [26] is a stepwise analysis procedure which adds or deletes one feature variable at each step. Details are given by Welch *et al.* [15]. Once this selection process is complete, the test feature vectors are classified into one of the three cloud classes: stratocumulus, cumulus, or cirrus.

Ten stratocumulus scenes, six cumulus scenes, and eight cirrus scenes are randomly selected for training the classifier. All of the 20 subregions within each scene are used as training data. Likewise, all of the subregions among the remaining five stratocumulus scenes, four cumulus scenes, and four cirrus scenes are used as holdouts to test the classifier.

Table I shows the classification results for textural measures defined by pixel separation $d = 1$. In this case, 66/100 of the stratocumulus (Sc) are classified correctly, with 23/100 stratocumulus misclassified as cumulus (Cu) and 11/100 as cirrus (Ci). Regions of stratocumulus breakup tend to be misclassified as cumulus, and regions of solid cloud cover tend to be misclassified as cirrus. Likewise, 75/80 of cumulus are classified correctly, with none misclassified as stratocumulus and 5/80 misclassified as cirrus. Finally, 77/80 cirrus are classified correctly, with two cirrus misclassified as stratocumulus and 1/80 as cumulus. The overall classification accuracy is 84%.

TABLE I
CLASSIFICATI●N ACCURACY FOR 512 × 512 PIXEL SAMPLES USING 67% OF THE SAMPLE FOR TRAINING

	Gray Level Difference Method		
Actual Classified	Sc	Cu	Ci
Sc	66	23	11
Cu	0	75	5
Ci	2	1	77
Overall Accuracy	84%		
Ranked	Standard Deviation		
Most	Local Homogeneity		
Important	Number of Cloud Pairs		
Textural	Contrast		
Features	Angular Second Moment		
Used by the	Entropy		
Classifier			

TABLE II
CLASSIFICATI●N ACCURACY FOR 512 × 512 PIXEL SAMPLES NEURAL NETWORK APPROACH

	Four Layers—1●% Sampling Set		
Actual Classified	Sc	Cu	Ci
Sc	79	0	21
Cu	4	56	20
Ci	1	0	79
Overall Accuracy	83%		
	Four Layers—20% Sampling Set		
Actual Classified	Sc	Cu	Ci
Sc	92	2	6
Cu	1	54	5
Ci	0	3	77
Overall Accuracy	93%		
	Three Layers—20% Sampling Set		
Actual Classified	Sc	Cu	Ci
Sc	93	1	6
Cu	1	55	4
Ci	●	18	62
Overall Accuracy	88%		

The most important variables used in the final step of the classifier are listed in Table I. Notably missing from this list are the measures of mean, and the third (shade), and fourth (prominence) moments of the difference vector density function.

Welch *et al.* [15] showed cloud classification results for the gray level co-occurrence matrix approach. They found that results for pixel separations of $d = 1, 2, 4$, and 8 were essentially equivalent. However, at larger pixel separations, classification accuracy was found to decrease. They also examined classification results using multiple pixel separations, and found only a slight improvement in overall accuracy. Therefore the present study considers only results for pixel separation $d = 1$.

B. Neural Network

In supervised training used in this study, the neural network monitors performance internally. An error signal is generated by the system when the neural network produces an incorrect response. This error signal is fed back to the system, and the correct response is produced after a number of iterations.

In the present study, the Neural Ware, Inc., Professional II software package is used. The architecture consists of four layers with feedforward and backpropagation, as described below.

1) The *input layer* consists of the nine textural features defined in Section IV and one bias node. The function of the bias node is to allow the network to respond actively to a signal with small values of each input. The absence of input signals is significant.

2) The *first hidden layer* consists of 40 components. There is no simple method of determining the minimum number of hidden nodes for a given problem. If the system has fewer nodes than are required for the specific problem, the iteration process can become "stuck" at a local minimum of the error function. We find that 40 components are sufficient for the present problem.

3) The second hidden layer consists of 20 components. Fewer than 20 nodes do not extract enough information from the previous layer to separate the classes adequately. Too many nodes slow down the learning phase without improving accuracy.

4) The *output layer* consists of three components; namely, the classifications of stratocumulus, cumulus, and cirrus cloudiness.

In the first classification scheme, three randomly selected subregions are selected each, from ten randomly selected stratocumulus scenes, from six randomly selected cumulus scenes, and from eight randomly selected cirrus scenes. Therefore only 72 (10%) of the total 740 subregions are used for training the classifier. Only data from the remaining 13 independent scenes are used for testing the classifier, consisting of 260 subregions. An overall classification accuracy of 83% is achieved. This is approximately the same accuracy obtained using linear stepwise discriminant analysis. As shown in Table II, 79/100 of the stratocumulus are classified correctly, with 21/100 misclassified as cirrus. Likewise, 56/80 cumulus are classified correctly, with 4/80 misclassified as Sc and 20/80 as Ci. Finally, nearly all of the cirrus are classified correctly.

In order to estimate the precision of the estimated classification accuracy of the neural network approach, a sample of 150 is randomly chosen from the 260 total test sample subregions and is tested with the NN classifier. The process is repeated six times with different randomly selected samples, as shown in Table III. The estimate of the classification accuracy is computed for each of the six groups of samples, and the mean and standard deviation of these estimates are computed. The mean of these estimates for the 10% sample size is 82%, with a standard deviation of 1.4%.

Increasing the size of the training set to 20% of the total

TABLE III
MEAN AND STANDARD DEVIATION OF THE CLASSIFICATION ACCURACIES USING NNW AND KNN CLASSIFIERS. BASED ON SETS OF SIX RANDOM SAMPLES OF 150 EACH FROM THE TEST DATA.

	% Accuracy with 10% Training Data		% Accuracy with 20% Training Data	
Classifier Type	Sample Set No.	% Accuracy	Sample Set No.	% Accuracy
NNW	1	80.00	1	92.67
	2	84.00	2	93.30
	3	82.67	3	94.00
	4	81.33	4	94.67
	5	81.33	5	94.00
	6	82.69	6	94.00
	Mean = 82.00%	S.D. = 1.40%	Mean = 94.00%	S.D. = 0.70%

	% Accuracy with 10% Training Data		% Accuracy with 20% Training Data	
Classifier Type	Sample Set No.	% Accuracy	Sample Set No.	% Accuracy
KNN	1	76.00	1	83.30
	2	71.30	2	84.00
	3	68.67	3	83.30
	4	70.67	4	84.67
	5	70.00	5	84.00
	6	69.30	6	86.67
	Mean = 71.00%	S.D. = 2.63%	Mean = 84.32%	S.D. = 1.30%

subregions, the overall classification accuracy increases to 93%. Both Sc and Cu show significant increases in classification accuracy. Estimated accuracy determined from the six sets of randomly selected samples is 94%, with a standard deviation of 0.7%, as shown in Table III. Finally, it is notable that the nonlinear, nonparametric neural network approach achieves classification accuracies much higher than those obtained using the linear stepwise discriminant approach, while utilizing a much smaller set of training data.

An estimate of theoretical classification accuracy often is obtained using the Monte Carlo approach. Recently, Ebert [27] and Garand [28] used this approach to estimate the accuracy of their cloud classification schemes. However, these approaches are based upon the assumption of multinormality of the distributions and are inappropriate for the present nonparametric approach. Furthermore, it is widely known that cloud textural features do not have normal distributions.

A more theoretically sound approach is the bootstrap method [29]-[31], which treats the simulation on a nonparametric basis. Welch *et al.* [15] computed cloud classification bootstrap estimates of correct classification probability and standard error for this dataset. Standard error estimates of 6-7% were obtained and can be assumed for the present study. Due to the large computing requirements in training the neural network, it was not feasible to apply the bootstrap approach in the present study.

Fig. 3 shows the significance of each feature vector to classification of each of the three cloud types for the 20% training set. The nine texture features defined in Section IV are listed along the ordinate. The abscissa shows the computed sum over all paths of the products or weights along a path for each path connecting the given texture feature to the given cloud type.

The justification of this measure is as follows. If the model were linearized, then the current multilayer model would reduce to a perceptron model in which the weight connecting the texture feature in the input layer to the cloud type in the output would be given by the values along the abscissa. Large values of these "linearized weights" correspond to texture features discriminating between cloud types. That is, large positive values indicate a strong positive association between the feature and the cloud type. In contrast, large negative values correspond to a negative association. For example, for the cirrus case shown in Fig. 3(c), a large negative weight for homogeneity strongly indicates this is *not* cirrus cloudiness.

A small positive or negative value indicates that this texture feature is not significant in identifying the cloud type. Clearly, the third (shade) and fourth (prominence) moments of the difference vector density function are not important texture features for cloud classification. Similar results are obtained for the 10% sampling sets.

Table II also shows classification accuracies derived from the three-layer neural network architecture for the 20% sample set. Both Sc and Cu show classification accuracies comparable to those obtained with the four-layer architecture. However, the three-layer results show a significant decrease in Ci classification accuracy, with 18/80 Ci misclassified as Cu. Overall accuracy drops from 93 to 88%. Estimated accuracy determined from the six sets of randomly selected samples is 88%, with a standard deviation of 0.94%. Clearly, for an equal number of sampling sets, the four-layer architecture produces higher classification accuracies.

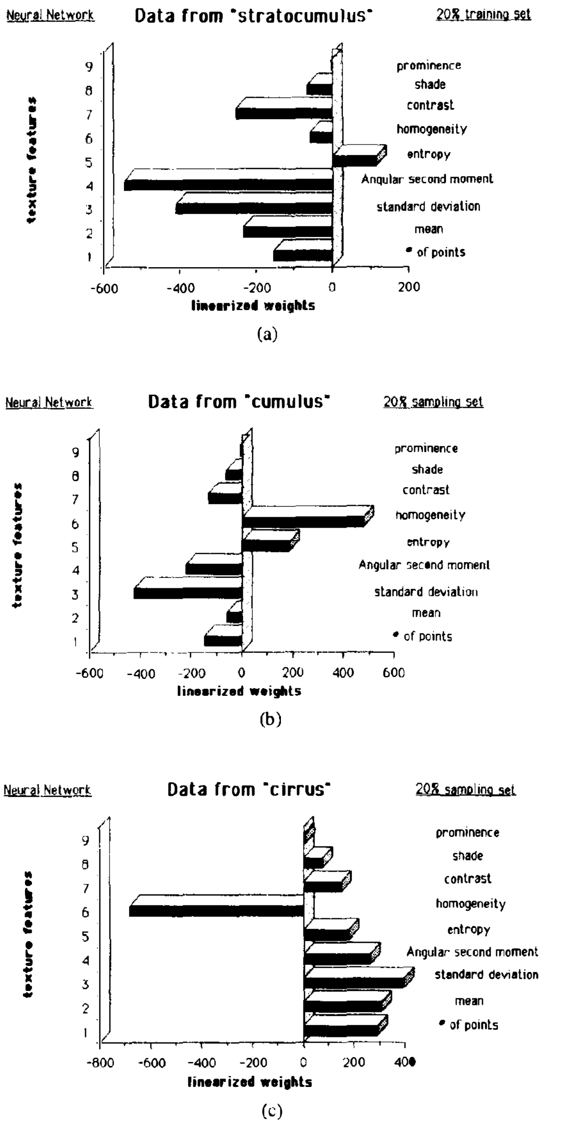


Fig. 3. Measurements of the significance of textural features derived from linearized neural network weights. Column 2 shows the linearized weight of each of the nine texture features shown in column 1.

C. *K-Nearest Neighbor*

The neural network approach is a nonlinear, nonparametric technique, while the discriminant analysis approach is linear. In order to attain a comparison of classification accuracy to another nonlinear, nonparametric approach, the *K*-nearest neighbor technique is applied.

First, three subregions are randomly chosen from ten stratocumulus, six cumulus, and eight cirrus scenes as training data. This represents 10% of the total subregions used as training data. Similar random selections of subregions are used to obtain a 20% training data. Table IV shows the classification accuracies for 10 and 20% training data. Overall classification accuracies of 73 and 85% are obtained, respectively.

TABLE IV
CLASSIFICATION ACCURACY FOR 512 × 512 PIXEL SAMPLES
K-NEAREST NEIGHBORS APPROACH

	10% Sampling Set		
Actual Classified	Sc	Cu	Ci
Sc	65	12	23
Cu	10	53	17
Ci	4	4	72
Overall Accuracy	73%		
	20% Sampling Set		
Actual Classified	Sc	Cu	Ci
Sc	81	6	13
Cu	0	48	12
Ci	4	1	75
Overall Accuracy	85%		

In order to estimate the accuracy of the KNN approach, a sample of 150 is randomly chosen from the 260 total test sample subregions and is tested with the KNN classifier. The process is repeated six times with different randomly selected samples. The mean classification accuracy is computed for each of the six groups of samples. Finally, the mean and standard deviation of these means are computed. For 10 and 20% training samples, the estimated mean accuracy of the classifier is 71 and 84.32%, respectively. The corresponding standard deviations are 2.63 and 1.3%, respectively. Clearly, the neural network approach yields an overall classification accuracy several percent higher than that obtained using KNN. Another important advantage of the neural network approach is that it provides information concerning the relative significance of the components of the input feature vector. Other nonparametric classifiers do not yield information of this kind.

VII. CONCLUSIONS

The cloud validation results of Stowe *et al.* [4] and Parker and Wielicki [6] clearly demonstrate that standard algorithms do not adequately retrieve cloud type and cloud cover. Recent results show that spatial information in the form of textural measures can greatly aid in this process [15], [27], [28]. However, the application of spatial pattern recognition techniques to cloud identification and analysis remain in their infancy. There is little information available as to the accuracies which can be attained using spatial features or in the choice of an optimal feature set. The present paper demonstrates that, with high spatial resolution data, very high cloud classification accuracies can be obtained, especially in conjunction with a nonparametric neural network classifier. The issue of textural classification accuracy as a function of spatial resolution is considered separately [32].

This study finds that a texture-based neural network classifier using only single-channel visible data can achieve overall cloud identification accuracy of 93%. Cirrus can be distinguished from low-boundary layer cloudiness with an accuracy of 96%, without the use of an infrared channel. Stratocumulus is retrieved with an accuracy of 92%, and cumulus is worst, with an accuracy of 90%. It is interesting to note that the use of the neural network has not improved cirrus classification accuracy. Rather, its main effect is in the improvement of stratocumulus classification. The neural network is better able to discriminate between stratocumulus and cumulus coordinates than is the standard stepwise discriminant analysis approach. The nonparametric KNN approach provided an overall accuracy of 85%, and the linear discriminant analysis provided an overall accuracy of 84%. The two nonparametric classifiers are able to obtain their high accuracies using only 20% of the database used for training. In contrast, the linear discriminant analysis procedure uses 67% of the database for training. For the same number of sampling sets, the four-layer neural network architecture produces higher cloud classification accuracies than does a three-layer architecture.

It is clear that the nonlinear, nonparametric neural network classifier provides the highest cloud classification accuracy of the three approaches, and is able to do so using a relatively small training set. It also yields important information concerning the relative significance of the components of the input feature vector. Other nonparametric approaches also require small training sets, but they do not provide significance estimates of the components of the input feature vector.

In a serial implementation, the training of the neural network may be long compared to other methods. The present study required 120 000 iterations of backpropagation to train the NNW. The choice of the learning rate η in (4) affects the speed of convergence. A small value of η causes a slow learning rate, but a large value of η causes an erratic oscillating behavior near a sharp change in the global error (3). The present study has not examined the speed of convergence for different values of η. However, texture-based cloud classification is rapid once training is completed. It should be noted that at the time of publication, a new NNW architecture has been developed which requires only 50 backpropagation iterations to train the same data. This is a savings of over two orders of magnitude in training time, with only a slight decrease in the overall accuracy. Details will be presented in a separate publication.

It should also be noted that the present study does not attempt the linkage between observed radiances and derived cloud parameters. For example, cloud optical thickness can be retrieved from reflected solar radiation measurements [33], [34], and cloud phase and effective particle size can be retrieved from the near-infrared channels [35]. Cloud parameters of climatological significance include cloud top and base heights, cloud cover for each cloud type, cloud field morphology, and cloud optical thickness, phase, particle size, and emissivity. The retrieval of these cloud properties is beyond the scope of the present investigation, which is focused upon the accurate retrieval of cloud type. Furthermore, many of these derived cloud products require infrared and near-infrared channels which the present data set does not contain.

In addition to the ISCCP effort to accurately discriminate cloud type, the Earth Observing System (EOS) and the Cloud and the Earth's Radiant Energy System (CERES) Science Team has identified three additional reasons for accurate scene identification. First, categorization into cloud-scene types is necessary for selection of the appropriate bidirectional reflectance angular model for converting radiance at the instrument to flux leaving the top of the atmosphere. Second, categorization into cloud-scene types is necessary for proper averaging, particularly in correcting for the solar zenith angle dependence of albedo. Finally, and perhaps more important, is the fact that categorization into cloud-scene types provides the capability of directly assessing changes in clear-sky albedo and long-wave flux.

The present results clearly establish the high spatial information content of satellite imagery for cloud identification. It is important to obtain the highest possible classification accuracy for the global monitoring of planet Earth. The combination of textural-based features in a neural network classifier certainly should be considered in future cloud retrieval algorithms, especially over polar regions.

Acknowledgment

Textural computations were performed on the South Dakota School of Mines and Technology CYBER computer, and neural network computations were performed on an IBM PC using the Neural Works, Professional II of Neural Ware Inc. software package.

Appreciation is extended to J. Robinson and S. Palmer for typing this paper, and to D. Adams for the photographic work. Thanks are also due to M. Navar for computational assistance.

References

[1] N. A. Hughes, "Global cloud climatologies: A historical review," *J. Climate Appl. Meteor.*, vol. 23, pp. 724–751, 1984.

[2] A. H. Goodman and A. Henderson-Sellers, "Cloud detection and analysis: A review of recent progress," *Atmos. Res.*, vol. 21, pp. 203–228, 1988.

[3] W. B. Rossow, "Measuring cloud properties from space: A review," *J. Climate*, vol. 2, pp. 201–213, 1989.

[4] L. L. Stowe, H. Y. M. Yeh, T. F. Eck, C. G. Wellemeyer, H. L. Kyle, and the Nimbus-7 Cloud Data Processing Team, "Nimbus-7 global cloud climatology. Part I: First year results," *J. Climate*, vol. *J. Climate*, vol. 2, pp. 671–709, 1989.

[5] J. Hansen *et al.*, "Climate impact of increasing atmospheric carbon dioxide," *Science*, vol. 213, pp. 957–966, 1981.

[6] L. Parker and B. A. Wielicki, "Comparison of satellite based cloud retrieval methods for cirrus and stratocumulus," presented at the FIRE Workshop, Monterey, CA, July 1989.

[7] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193–202, 1980.

[8] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks*, vol. 1, pp. 75–89, 1988.

[9] D. Rumelhart, G. Hinton, and R. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, D. Rumelhart and J. McLelland, Eds. Cambridge, MA: MIT Press, 1986, pp. 318–362.

[10] E. Collins, S. Ghosh, and C. Scofield, *Risk Analysis, DARPA Neural Network Study.* Fairfax, VA: AFCEA Int. Press, 1988.

[11] G. N. Reeke, Jr. and G. M. Edelman, "Selective neural networks and their implications for recognition automata," *Int. J. Supercomput. Appl.*, vol. 1, pp. 44-69, 1987.

[12] R. M. Haralick, K. S. Shammugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610-621, 1973.

[13] K. S. Kuo, R. M. Welch, and S. K. Sengupta, "Structural and textural characteristics of cirrus clouds observed using high spatial resolution Landsat imagery," *J. Appl. Meteor.*, vol. 27, pp. 1242-1260, 1988.

[14] R. M. Welch, S. K. Sengupta, and K. S. Kuo, "Marine stratocumulus cloud fields off the coast of southern California observed using Landsat imagery. Part II: Textural analysis," *J. Appl. Meteor.*, vol. 27, pp. 362-378, 1988.

[15] R. M. Welch, S. K. Sengupta, and D. W. Chen, "Cloud field classification based upon high spatial resolution textural features. Part I: Gray level cooccurrence matrix approach," *J. Geophys. Res.*, vol. 93, pp. 12 663-12 681, 1988.

[16] M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 118-125, Jan. 1986.

[17] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 2269-2285, 1976.

[18] D. W. Chen, S. K. Sengupta, and R. M. Welch, "Cloud field classification based upon high spatial resolution textural features. Part 2: Simplified vector approaches," *J. Geophys. Res.*, to be published.

[19] R. Hecht-Nielsen, "Nearest matched filter classification of spatiotemporal patterns," Hecht-Nielsen Neuro-Comput. Corp., June 1986.

[20] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci.*, vol. 74, pp. 2554-2558, 1984.

[21] G. E. Hinton, T. J. Sejnowski, and S. D. Ackley, "Boltzmann machine," Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-84-119, 1984.

[22] G. G. Lorentz, "The 13th problem of Hilbert," in *Mathematical Developments Arising from Hilbert Problems*, F. E. Browder, Ed. Providence, RI: Amer. Math. Soc., 1976.

[23] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973, 482 pp.

[24] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

[25] R. I. Jennrich, "Stepwise discriminant analysis," in *Statistical Meth●ds for Digital Computers*, K. Euslein, A. Ralston, and H. S. Wilf, Eds. New York: Wiley, 1977. 454 pp.

[26] A. A. Afifi and V. S. P. Azen, *Computer Aided Multivariate Analysis.* Lifetime Learning Publ., 1984, 458 pp.

[27] E. Ebert, "A pattern recognition technique for distinguishing surface and cloud types in the polar regions," *J. Climate Appl. Mete●r.*, vol. 26, pp. 1412-1427, 1987.

[28] L. Garand, "Automated recognition of oceanic cloud pattern. Part I: Methodology and application to cloud climatology," *J. Climate*, vol. 1, pp. 20-39, 1988.

[29] B. Effron, "The jack-knife, the bootstrap and other resampling plans," in *CBMS-NSF Ser. SIAM*, Pittsburgh, PA, 1982.

[30] S. Chatterjee and S. Chatterjee, "Estimation of misclassification probabilities by bootstrap methods," *Commun. Statist. Simulation Comput.*, vol. 12, no. 6, pp. 645-656, 1983.

[31] A. K. Jain, R. Dubis, and C. C. Chen, "Bootstrap techniques for error estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, 1987.

[32] R. M. Welch, M. S. Navar, and S. K. Sengupta, "The effect of spatial resolution upon texture-based cloud field classifications," *J. Geophys. Res.*, Oct. 1989.

[33] A. Arking and J. D. Childs, "Retrieval of cloud-cover parameters from multispectral satellite measurements," *J. Climate Appl. Meteor.*, vol. 24, pp. 322-333, 1985.

[34] M. D. King, "Determination of the scaled optical thickness of clouds from reflected solar radiation measurements," *J. Atmos. Sci.*, vol. 44, pp. 1734-1751, 1987.

[35] B. A. Wielicki *et al.*, "The 27-28 October 1986 FIRE IFO case study: Comparison of radiative transfer theory with observations by satellite and aircraft," *Mon. Wea. Rev.*, to be published.

Jonathan Lee received the B.S. degree in computer science from the South Dakota School of Mines and Technology, Rapid City, where he is currently working on the M.S. degree in computer science, with a concentration in neural network, distributed/parallel system, and human-machine interfacing. He is a Software Engineer at the MCS Group, Inc., Rapid City.

Mr. Lee is an associate member of the Sigma Xi Scientific Research Society of America.

*

Ronald C. Weger was born on February 22, 1941. In 1962 he received the B.S. degree in mathematics from William Jewell College, Liberty, MO. He then obtained the M.S. degree in mathematics in 1964 from the University of Illinois, Urbana. In 1967, at the same institution, he earned the Ph.D. degree in mathematics.

In 1967 he was appointed Assistant Professor at the South Dakota School of Mines and Technology, Rapid City, in the Mathematics Department, and was promoted to Associate Professor and then Professor in 1972 and 1976, respectively. He has special expertise in the area of computer science. In the summer of 1982 he developed a spreadsheet for VAX 780 for Martin Associates. Then during 1983-1984 he worked on a BASIC Compiler for MC68000 for the same company. Since 1984 he has developed several compilers for the MC68000 Microprocessor in various PC environments.

Dr. Weger is a member of the Association for Computing Machinery, the Mathematical Association of America, Sigma Xi, and Pi Mu Epsilon. He received the Standard Oil Good Teaching Award in 1971.

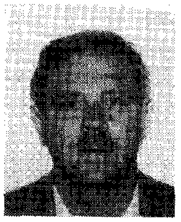
*

Sailes K. Sengupta was born in Bankura, India. He received the B.S. and M.S. degrees in mathematics from Calcutta University, and the Ph.D. degree in statistics from the University of California, Berkeley.

He has been an Assistant Professor of Mathematics at the University of Missouri, Kansas City, and Associate Professor of Mathematical Sciences at the South Dakota School of Mines and Technology, Rapid City, where since 1981 he has been a Professor of Mathematics and Computer Science. He has several publications in the areas of texture-based classification and pattern recognition of cloud fields. His research interests include image processing, pattern recognition, expert systems and neural computing.

Dr. Sengupta is a member of ACM, Sigma Xi, and SPIE.

*



Ronald M. Welch (M'88) received the B.S. and M.A. degrees in physics in 1965 from California State University, Long Beach. In 1971 he earned the Ph.D. degree in physics, and the Ph.D. degree in meteorology in 1976 from the University of Utah, Salt Lake City.

From 1976-1978 he was a Research Associate in the Department of Atmospheric Sciences at Colorado State University, and at the Institute for Meteorology at Johannes Gutenberg University, Mainz, West Germany, during 1978-1981. He was Associate Professor at Old Dominion University during 1981-1982. Since 1982 he has been Associate Professor and Head, Data Acquisition and Analysis Group, Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, where he was promoted to Senior Scientist in 1987 and to Professor in 1990. Presently, he is at the Naval ●ceanographic and Atmospheric Research Laboratory in Monterey, CA, under a UCAR Senior Fellowship.

Dr. Welch serves on the Eos HIRIS, ITIR, and CERES Instrument Science Teams, and on the E●sDIS Advisory Panel.