Application of Multilayer Feedforward Neural Networks to Precipitation Cell-Top Altitude Estimation

Michelle S. Spina, Michael J. Schwartz, David H. Staelin, *Fellow, IEEE*, and Albin J. Gasiewski, *Senior Member, IEEE*

***Abstract*—The use of passive 118-GHz O_2 observations of rain cells for precipitation cell-top altitude estimation is demonstrated by using a multilayer feedforward neural network retrieval system. Rain cell observations at 118 GHz were compared with estimates of the cell-top altitude obtained by optical stereoscopy. The observations were made with 2–4-km horizontal spatial resolution by using the millimeter-wave temperature sounder (MTS) scanning spectrometer aboard the NASA ER-2 research aircraft during the Genesis of Atlantic Lows Experiment (GALE) and the Cooperative Huntsville Meteorological Experiment (COHMEX) in 1986. The neural network estimator applied to MTS spectral differences between clouds, and nearby clear air yielded an rms discrepancy of 1.76 km for a combined cumulus, mature, and dissipating cell set and 1.44 km for the cumulus-only set. An improvement in rms discrepancy to 1.36 km was achieved by including additional MTS information on the absolute atmospheric temperature profile. An incremental method for training neural networks was developed that yielded robust results, despite the use of as few as 56 training spectra. Comparison of these results with a nonlinear statistical estimator shows that superior results can be obtained with a neural network retrieval system. Imagery of estimated cell-top altitudes was created from 118-GHz spectral imagery gathered from CAMEX, September through October 1993, and from cyclone Oliver, February 7, 1993.**

***Index Terms*—Microwave remote sensing, microwave spectra 118 GHz, neural network, precipitation estimation, rain cell-top altitude.**

I. Introduction

CLOUD and precipitation parameters can be remotely sensed from brightness temperature measurements made at millimeter wavelengths from space. The potential of passive microwave remote sensing of rainfall rate has been discussed by Wilheit [20] and retrievals of rainfall rate have been demonstrated from 19.35-GHz passive data [21] and from combined 18- and 37-GHz passive data [17]. Frequencies within 2.5 GHz of the 118.750-GHz O_2 line have been shown to support retrievals of precipitation parameters, in particular, cell-top altitude and area [8]. It has been separately shown that thunderstorm cloud height is statistically related to rainfall rate, although the relationship is strongly influenced by climatological region [1]. In addition, correlations between the maximum cell-top altitude, and both the total rainfall volume and the maximum rainfall volume rate have been revealed [6]. Information provided by independent cell-top altitude estimates would be beneficial over highly glaciated cells, where rainfall-rate retrievals using only 19- and 37-GHz frequencies can be compromised by wet soil or a cloud layer containing strongly scattering ice.

Cell-top altitude retrievals have been demonstrated using passive measurements of the infrared radiance emitted at the cloud top [16]. However, because of scattering and absorption, passive infrared observations generally cannot probe beneath nonprecipitating cloud canopies. At microwave frequencies, there is significantly less extinction of the radiation, allowing for the direct probing of the larger precipitation particles located beneath any optically obscuring cloud canopies.

The embedding of cell-top altitude information in 118-GHz observations occurs through two mechanisms. First, statistical dependences between the cell-top altitude and the total water content and phase (liquid or ice) produce statistical dependences between the cell-top altitude and brightness perturbations in the transparent 118-GHz channels. For example, consider cells with tops below the freezing level: Higher cell-top altitudes are associated with increased precipitation. The increase in water density causes increased absorption, which produces a decrease in brightness temperatures. In the case of cells with tops above the freezing level, an increased quantity of ice exists in the cell top. The presence of ice causes strong scattering of the cold cosmic background radiation, which produces large negative perturbations in brightness temperature.

Second, the altitude distribution of atmospheric water can be probed by the virtue of the successively higher peaking altitudes of the 118-GHz clear-air temperature weighting functions. Low altitude cells produce little or no signature in the opaque 118-GHz channels, while high altitude cells cause perturbations in all 118-GHz channels. The 118-GHz clear-air weighting functions peak at altitudes ranging from the surface, for transparent frequencies located 2.5 GHz from the 118.750-GHz line center frequency, to more than 35 km, which is far above the tops of most clouds, for frequencies at the line center. Clear-air weighting functions for the millimeter-wave temperature sounder (MTS) are shown in [9].

This paper evaluates retrievals of cell-top altitude by using multilayer, feedforward neural networks operating on high-

Manuscript received January 19, 1995; revised January 19, 1996. This work was performed at the Research Laboratory of Electronics, Massachusetts Institute of Technology, and was supported by NASA Grants NAG5-10 and NAG5-2545, and NASA Contract NAS5-30791.

M. S. Spina, M. J. Schwartz, and D. H. Staelin are with the Massachusetts Institute of Technology, Cambridge, MA 02139-3594 USA (e-mail: staelin@ll.mit.edu).

A. J. Gasiewski is with the Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

Publisher Item Identifier S 0196-2892(98)00142-9.

resolution, passive 118-GHz multichannel precipitation cell imagery. The ability of these networks to approximate complex mathematical relations has been shown in the literature [12], [13], and their use in remote sensing applications has been demonstrated [2]–[4], [18], [19]. Neural networks estimate precipitating cell-top altitudes from 118-GHz spectral data by capturing both the complex statistical nature of the spectral data and the nonlinear relationship that exists between 118-GHz spectral emissions and cell-top altitudes. An optimal mapping of 118-GHz spectral data to cell-top altitude is accomplished by training a multilayer, feedforward neural network by using a set of examples that characterize the statistical complexity of the estimation problem. The cell-top altitude estimate produced by the neural network is optimized with respect to a mean-squared-error criterion using the backpropagation algorithm to adjust the parameters of the neural network. An incremental training algorithm was also developed to allow the incorporation of weakly correlated data into the network without substantially increasing the complexity and size of the network.

II. Passive Microwave Remote Sensing of Cell-Top Altitude

A. Retrieval of Cell-Top Altitude from 118.75-GHz Spectra

Precipitating clouds perturb retrieved temperature profiles of the atmosphere for two reasons. First, clouds below the freezing level absorb upward propagating radiation and, therefore, usually decrease the observed brightness temperature. Second, larger ice particles in convective clouds above the freezing level can strongly reflect the cold cosmic background radiation, causing large negative perturbations in brightness temperature. It has been demonstrated by Gasiewski *et al.* [9] that cloud and precipitation properties can be remotely sensed from brightness temperature measurements made at frequencies within 2 GHz of the 118.75-GHz O_2 resonance. The more opaque frequencies respond only to the highest cloud tops and, thus, provide a means for measuring cloud altitude. The retrieval system described in this paper has been developed specifically for precipitation cells since only such cells exhibit ice densities high enough to provide a reflection signature that is relatively independent of the cloud thickness.

To develop a cell-top altitude retrieval system a collection of 279 independent near-nadir brightness temperature spectra of precipitation cell cores was compiled. The collection consists of spectra produced by the MIT MTS scanning spectrometer aboard the NASA ER-2 aircraft during GALE [5], in which 19 spectra were collected, and the COHMEX [22], in which 260 spectra were collected. These observations represent precipitation observed during seven winter and 14 summer aircraft flights, respectively. Except for a single test flight and ER-2 ferry flight, all flights during GALE were out of Patrick Air Force Base, Florida. The objective of GALE was to observe stratiform precipitation over the southeastern coastal region of the United States. During COHMEX, all flights originated from NASA Wallops Island Research Center, Wallops Island, VA. The objective was to observe summer mid-latitude convective precipitation over the Huntsville, AL, area and in the vicinity of Wallops Island.

A thorough description of the instrument, aircraft flights, and data calibration is given in Gasiewski *et al.* [9]. In summary, the 118-GHz spectrometer is a double-sideband, super-heterodyne receiver with a 118.75-GHz local oscillator and eight i.f. filters of 200-MHz width, which together cover the band 400–2000 MHz. The 7.5° beamwidth antenna has a flat rotating mirror that deflects the beam 90° and scans it through nadir in 14 7.2° steps, plus two steps viewing hot and ambient-temperature calibration loads, the former being 45–85-K warmer. For cloud tops near 10-km altitude and the aircraft near 20 km, the horizontal spatial resolution is 1.3 km at nadir, increasing to 1.8 × 2.6 km at the edge of the swath; these distances are doubled at the terrestrial surface.

The measured 118-GHz spectrum for a precipitating cell is denoted by an eight-dimensional (8-D) vector of brightness temperature observations

$$\overline{T}_{Bi} = \begin{bmatrix} T_{B(118.75\pm0.50)i} \\ T_{B(118.75\pm0.66)i} \\ T_{B(118.75\pm0.84)i} \\ T_{B(118.75\pm1.04)i} \\ T_{B(118.75\pm1.26)i} \\ T_{B(118.75\pm1.47)i} \\ T_{B(118.75\pm1.67)i} \\ T_{B(118.75\pm1.90)i} \end{bmatrix}$$

where the subscripts indicate the local oscillator and central sideband frequencies (GHz) for the MTS, arranged in order of decreasing opacity. For each rain-cell spectrum ($\overline{T}_{Bi}$), a corresponding clear-air reference spectrum ($\overline{T}_{Br}$) was estimated from MTS observations in the vicinity of the cell. A delta brightness spectrum ($\Delta\overline{T}_{Bi}$) was determined as the difference between the cell spectrum and the clear-air reference spectrum

$$\Delta\overline{T}_{Bi} = \overline{T}_{Bi} - \overline{T}_{Br}.$$

Delta brightness spectra were used in this retrieval system rather than the absolute brightness spectra to ensure that any fluctuations in the baseline brightness spectra among cell observations, due to fluctuations in the ambient atmospheric temperature profile, were removed. Typical delta brightness spectra and clear-air reference spectra were presented by Gasiewski *et al.* [8]. The largest delta brightnesses vary from −40 to −170 K over the i.f. band 0.5–2 GHz, which is large compared to receiver sensitivities of 1 K.

The altitude of each cell-top was estimated by stereoscopy, using MTS video images (VHS color images through a 99° wide-angle lens) and the known altitude and air speed of the aircraft. The visually estimated cell-top altitude has associated rms errors of 1 km, due to uncertainties in aircraft speed relative to the cloud (±10%), aircraft altitude (±500 m), and time of passage of a particular feature of a cell-top through the video field of view (±2 s). The size of each cell was estimated by using the MTS spectral imagery and was defined as the distance along the aircraft flight track over which the MTS transparent channel brightness perturbation decreased to half its maximum value. For elongated rain cells, the geometric mean of the major and minor horizontal dimensions was used.

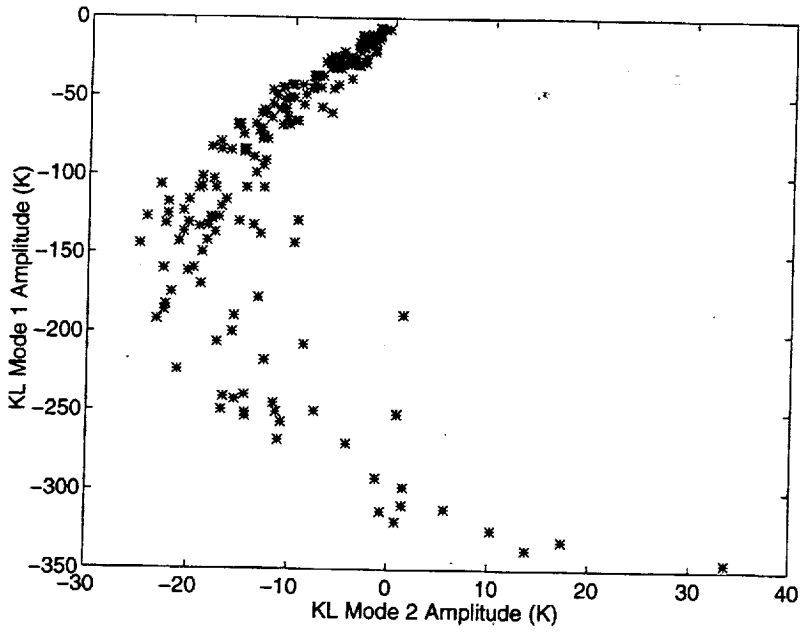


Fig. 1. Scatter plot of KLT mode 1 amplitude versus KLT mode 2 amplitude.

In addition to the collection of brightness temperature spectra and cell-top altitudes, observed cells used in this retrieval were classified as one of two types. Cells that appeared to be in their early stages of convection were designated "cumulus-type," while those exhibiting anvils were designated "mature-type." The cells were visually classified by using the MTS video imagery. The collection of cells ranges from 2 to 17 km in cell-top altitude and from 5 to 200 km in cell-size. It can be shown that the logarithm of cell size is linearly correlated with the cell-top altitude. This apparent relationship was exploited in the development of the neural network retrieval system.

B. Discussion of Previous Methods

A nonlinear statistical estimator operating on perturbation spectra ($\Delta\overline{T}_{Bi}$) was developed by Gasiewski and Staelin [8] to estimate cell-top altitude. The estimator consisted of an orthogonal Karhunen-Loève transformation (KLT) [23], followed by a rank reduction operation, a nonlinear operator, and a linear-statistical estimator. The KLT and rank reduction operations were used to reduce the complexity of the perturbation spectra by removing any redundancy that may exist between channels.

The KLT is performed by diagonalizing the covariance matrix of the perturbation spectra

$$\Re_{\Delta\overline{T}_B\Delta\overline{T}_B} = \Sigma' \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_8 \end{bmatrix} \Sigma$$

where $\Re_{\Delta\overline{T}_B\Delta\overline{T}_B}$ is the covariance matrix, Σ is the row-matrix consisting of the eigenvectors of $\Re_{\Delta\overline{T}_B\Delta\overline{T}_B}$, λ_i are the associated eigenvalues and also a measure of the variance of the ith component of the decomposed spectra, and ($'$) is the transpose operator. The KLT is performed by computing $k_i = \Sigma\Delta T_{Bi}$. The eight eigenvalues of the covariance matrix were found to be 7589.3, 65.7, 1.7, 0.3, 0.2, 0.2, 0.1, 0.1. Since the instrument sensitivity is 1 K, only the first and second KLT modes contain statistically useful information. These two most dominant KLT coefficients were then linearized with respect to the cell-top altitude, and a linear statistical estimator operating on these two linearized coefficients was then used to minimize the mean-squared error between the estimated and actual cell-top altitude.

This nonlinear KLT-based estimator produces better results with respect to the mean-squared-error criterion than a simple linear regression estimator for the cell-top altitude retrieval problem. However, the KLT would fully characterize the statistical behavior of the spectra only if they had a jointly Gaussian probability distribution. If this were the case, the KLT mode 1 and KLT mode 2 variables would exhibit an elliptically shaped joint probability distribution (pdf). As shown by Fig. 1, the joint pdf for KLT mode 1 and KLT mode 2 is clearly not elliptical. In addition, it can be seen from Fig. 1 that the KLT mode 1 variables are not Gaussian-distributed. This indicates that additional information exists in higher order statistics, and improvement in cell-top altitude retrievals may be achieved with other methods that can capture this information. We, therefore, might reasonably expect neural networks to outperform estimators that merely combine second-order regression methods with nonlinear transformations, which linearize only some of the physical relationships between radiances and environmental parameters and cannot fully adapt to the statistical complexities of the data.

III. Neural Network Retrieval System

A. Introduction to Multilayer Feedforward Neural Networks

Artificial neural networks, or simply neural nets, are mathematical models that attempt to achieve good performance through interconnections of simple computational nodes. A

node is a single-valued function of multiple variables that is computed in two steps. First, a weighted sum of the node inputs is computed, and a bias term is added. This yields the linear output of the node. For all but the final output nodes, this output is then passed through a nonlinear function $f(x) = \tanh(x)$, referred to as the activation function of the node. Linear activation functions $f(x) = x$ are used in the output nodes for the cell-top retrieval network.

A feedforward network consists of layers of these simple computational elements operating in parallel, with the set of node outputs in a given layer providing the inputs to each of the nodes in the following layer. With such a network topology, no element can provide input to itself or to any other element that affects its input signals. The system input variables comprise the input layer of the network and hidden layers are comprised of variables that are not directly accessible to the outside world (they are neither input nor output variables). In the cell-top retrieval system, the inputs to the neural network consists of the vector of eight delta brightness values ($\Delta\overline{T}_B$), and the output variable is the cell-top altitude. Given the neural network topology, the weights and biases are determined to minimize the mean-squared error between the desired and calculated output variables by using the backpropagation algorithm developed by Rumelhart *et al.* [15]. Since there are currently no design rules that select an optimal network topology for a specific application, the topology must be determined by experiment. The performance criterion used was the mean-squared error between the retrieved and actual cell-top altitudes for a data set distinct from that used for training the network.

To improve upon the backpropagation algorithm, momentum, and an adaptive learning rate was implemented. The use of momentum decreases the sensitivity to small details in the error surface, which helps prevent the network from converging to a local rather than global minimum in the error surface. This is accomplished by allowing the network to respond not only to the local gradient, but also to recent trends in the error surface. An adaptive learning rate was used to allow the network to train faster by determining an optimal learning rate for the local terrain in the error surface.

B. Development of the Neural-Network-Based Altitude Retrieval System

The mapping from delta brightness spectra ($\Delta\overline{T}_B$) to cell-top altitude was accomplished by training a neural network with a subset of the $\Delta\overline{T}_B$ and associated cell-top altitudes. The remaining subset of $\Delta\overline{T}_B$ spectra was used to validate the network's performance on input–output pairs not previously seen by the network. Four types of data sets were used to train four different neural network cell-top altitude estimators. The data sets were derived from a collection of 279 independent near-nadir brightness temperature spectra compiled during GALE and COHMEX. Prior to both mapping and training, the $\Delta\overline{T}_B$ were scaled and offset to provide a zero-mean data set with normalized peak-to-peak variations.

The first data set consisted of 176 $\Delta\overline{T}_B$ spectra from the full collection of observed cloud types (both cumulus and mature), one spectrum per cloud. Although the visible spectrum is highly sensitive to the cirrus anvils, the 118-GHz channels often are not [9], and therefore, optical estimates of cell-top altitude are higher than the retrieved 118-GHz cell-top altitudes. For this reason, the remaining data set limits the observations to cumulus cloud types, which typically do not display these cirrus shields. This reduced the data set from 176 spectra to 84 spectra. Data set two consisted of cumulus-only data with $\Delta\overline{T}_B$ as input. Data set three consisted of cumulus-only data with both $\Delta\overline{T}_B$ and clear-air reference spectra ($\overline{T}_{Br}$) as input. Data set four incorporated the logarithm (base 10) of cell-size in addition to $\Delta\overline{T}_B$ and $\overline{T}_{Br}$ as input to the neural network estimator. Table I summarizes the data sets used in the development of the neural network retrieval system.

TABLE I
SUMMARY OF DATA SETS USED IN DEVELOPMENT OF CELL-TOP ALTITUDE RETRIEVAL SYSTEM

Input Data	# Training Patterns	# Test Patterns
All Cloud Types $\Delta\overline{T}_B$	117	59
Cumulus-Only Clouds $\Delta\overline{T}_B$	56	28
Cumulus-Only Clouds $\Delta\overline{T}_B$, $\overline{T}_{Br}$	56	28
Cumulus-Only Clouds $\Delta\overline{T}_B$, $\overline{T}_{Br}$,log(cell-size)	56	28

The development of the neural network retrieval system involved the selection of the network attributes and the training algorithm. The attributes of the network that needed to be determined were the network model and topology. This research explored the use of multilayer, feedforward neural networks trained by the backpropagation algorithm. Therefore, only the network topology (number of hidden layers, number of nodes in each layer) had to be determined. Currently, there are no design rules that indicate an optimal network topology for a specific application. Therefore, the optimal topology was determined by experimentation. It has been shown that networks with one or two hidden layers and a sufficient number of nodes are capable of approximating any well-behaved function [7], [11], [13], and therefore, only such networks were investigated. Seven network topologies were trained with a data set consisting of 117 $\Delta\overline{T}_B$ spectra from the combined mature and cumulus cloud types. The performance of the networks was based on the resulting rms error of the training and test sets. The test set consisted of 59 $\Delta\overline{T}_B$ spectra taken from the same collection of data as the training set. Since the training and test sets are mutually exclusive, the test set results show how well the network can generalize given similar data that it has not been trained on.

First, networks with two hidden layers were investigated. A network with six nodes in the first hidden layer and four nodes in the second hidden layer was tested first. The number of nodes in each of the hidden layers was increased until the performance of the test set began to degrade. A single hidden layer network with four hidden nodes was then evaluated. The number of hidden layer nodes was again increased until the performance of the test set began to

TABLE II
COMPARISON OF RMS ERRORS (km) FROM DIFFERENT NEURAL NETWORK TOPOLOGIES

Description of Neural Network Topology	Training Set RMS Error	Test Set RMS Error	# of Training Epochs Required
2 hidden layers: 6 and 4 nodes	1.81	1.78	3000
2 hidden layers: 8 and 4 nodes	1.70	1.79	3500
2 hidden layers: 8 and 6 nodes	1.71	1.79	4000
1 hidden layer: 4 nodes	1.77	1.80	25000
1 hidden layer: 5 nodes	1.74	1.75	3500
1 hidden layer: 6 nodes	1.71	1.76	5000
1 hidden layer: 7 nodes	1.72	1.75	5000

degrade. The training of each network was stopped when the training error converged to a minimum value. For each network topology, the resulting training and test set rms errors were noted as well as the number of presentations of the full training set (epochs) that was required for the network to reach a minimum error. Table II summarizes the results from the topology comparisons.

From these topology experiments, it was determined that a network with one hidden layer and five nodes was the optimal network topology for this data set. This network topology yields acceptable rms errors and performs as well on the test set as on the training set. The simpler network was able to outperform the more complex network for this application for a variety of reasons. First, the relatively small size of the training set limited the size of the network. Networks with a large number of nodes require a very large training set to satisfactorily constrain the parameters of the network. Second, the training of the large networks may have stopped in a local minimum or very flat region of the error surface.

The effect of weight initialization was also investigated. The weights from the input nodes to the first hidden layer were initialized by using the Nguyen–Widrow initialization method. This method has been shown to initialize the weights such that the network is able to converge to its final rms error value more quickly [14]. The remaining weights were initialized with random values. Experiments verified that the initial value of the weights did not affect the final network performance.

C. Development of the Incremental Neural Network

One typical problem encountered when developing neural network systems is that of overfitting to the training data. This problem of overfitting occurred in the development of the neural network estimator for the reduced data set consisting of 56 cumulus-only clouds. When the additional inputs of $\overline{T}_{Br}$ and log (cell-size) were added to the system, the network acquired 45 new weights [(8 + 1 new inputs) ×5 hidden nodes] in addition to the 45 weights already present (8 inputs ×5 hidden nodes +5 hidden nodes ×1 output). These additional weights provided enough new degrees of freedom

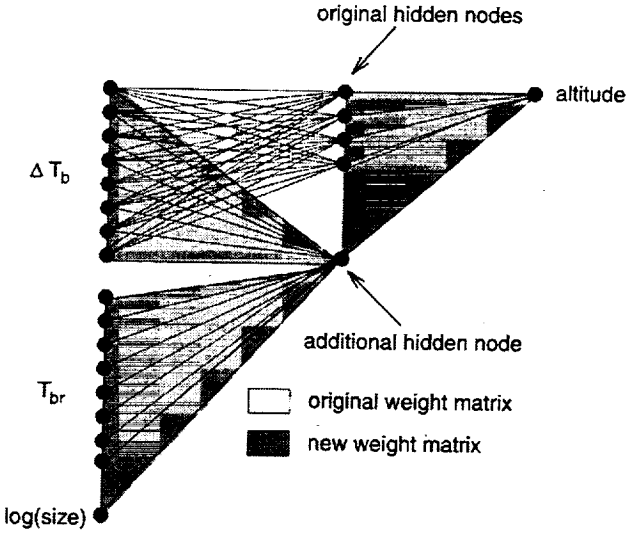


Fig. 2. Illustration of incremental neural network system.

that the network easily fit the training data, but the test set errors increased due to overfitting.

Although limiting the number of hidden nodes and limiting the number of epochs of training helped reduce overfitting, the resulting performance was not better than for systems with only the perturbation brightness temperature spectra ($\Delta\overline{T}_B$) as input. To accommodate the additional inputs [$\overline{T}_{Br}$ and log (cell-size)] without substantially increasing the size of the neural network, an "incremental neural network" training algorithm was developed.

When the clear-air reference spectra ($\overline{T}_{Br}$) and log (cell-size) were added as inputs to the neural network, the performance should exceed that given by the network with only $\Delta\overline{T}_B$ as input. To accomplish this, a network using $\Delta\overline{T}_B$ as input was initially trained by using one hidden layer with four nodes until the network rms error converged to a minimum. Then one additional hidden node was added, driven by the new inputs plus the original $\Delta\overline{T}_B$ inputs. The new inputs were not connected to the original four nodes and the weights and biases associated with the initially trained network were held constant, while the backpropagation algorithm was used to train the new weights and biases. The incremental neural network approach is illustrated graphically in Fig. 2. The original weight matrix is shown with a white background. The additional weight matrix is shown with a shaded background.

The topology illustrated in Fig. 2 has a total of 54 weights. If a fully connected network with five hidden nodes were used, the network would have a total of 90 weights. By dramatically decreasing the number of weights and using incremental training, the additional inputs could be successfully incorporated into the network without substantially increasing the size and complexity of the network. An investigation of whether or not this technique is applicable to other neural network applications would be interesting.

D. Retrieval Results

Table III shows the resulting rms altitude errors (km) for three retrieval methods and four data sets. The *a priori*

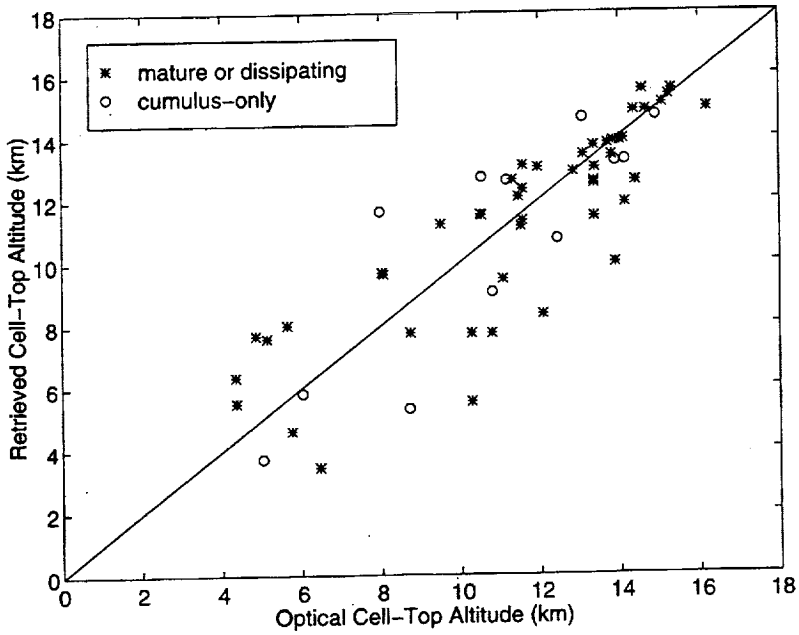


Fig. 3. Retrieved cell-top altitude versus cell-top altitude estimated by optical stereoscopy for the complete data set of cumulus, mature and dissipating cells (data set 1), using the neural network estimator.

TABLE III
COMPARISON OF RMS ERRORS (km) FOR CELL-TOP ALTITUDE ESTIMATORS

Data Set Used	A-Priori Variance	Linear Regression Divided	Linear Regression	Non-Linear Statistical Estimator	Neural Network Estimator
All Cloud Types ΔT_B	3.53	2.01	2.03	1.97	1.76
Cumulus-Only Clouds ΔT_B	3.53	2.27	1.82	1.63	1.44
Cumulus-Only Clouds ΔT_B, T_{B_r}	3.53	2.78	1.66	1.53	1.41
Cumulus-Only Clouds ΔT_B, T_{B_r}, log(cell-size)	3.53	2.56	1.58	1.50	1.36

variance in the full collection of altitude data is also listed. These three methods include simple linear regression, the nonlinear statistical estimator described previously, and neural networks using the optimal architectures that were previously determined. The error reported for the neural network estimator is the rms error associated with the validation set. The error reported for the linear-regression-divided estimator is the rms error associated with the validation set when the estimator was developed with the training set data only. The error reported for the linear regression estimator is the rms error associated with the entire data set. The nonlinear statistical estimator error shown was computed by Gasiewski *et al.* [8] and is also the rms error associated with the entire data set.

As illustrated by Table III, the neural network estimator outperforms the nonlinear statistical estimator by 10.7% for data set 1, 11.7% for data set 2, 7.8% for data set 3, and 9.3% for data set 4. Had the nonlinear statistical estimator been trained and tested on separate data, as was the neural network, its performance presumably would have been slightly degraded. Since the neural network retrievals yield 1.4-km rms discrepancies with stereoscopically determined altitudes and since the rms uncertainty of stereoscopy is roughly 1 km, as discussed earlier, the inferred rms accuracy of the 118-GHz retrieval is closer to 1 km. Values of cell-top altitude retrieved from the 118-GHz spectra for the complete data set (data set 1) are plotted against the optically estimated values in Fig. 3. The rms error for this retrieval was 1.76 km, with a correlation coefficient between the retrieved and optically estimated altitudes of 0.86. The bias toward the lower right side of the graph is due to the difference in the cell-top optical depth between the microwave and visible regions of the spectrum. As described previously, the visible spectrum is highly sensitive to cirrus anvils, while the 118-GHz channels are not. Therefore, the optical estimates of cell-top altitude are higher than the retrieved 118-GHz cell-top altitudes. Since cumulus cloud types typically do not display the cirrus shields, we expect that the retrieval results will be closer to the optically estimated values. The rms error of this reduced data set (data set 2) was reduced to 1.44 km, and the correlation coefficient between the retrieved and optically estimated altitudes was increased to 0.93.

E. Images of Retrieved Cell-Top Altitude

Cell-top altitude imagery was created from the output of the neural network estimator, from data collected during the Convection And Moisture Experiment (CAMEX) [10]. CAMEX was a multidisciplinary experiment designed to measure the three-dimensional (3-D) moisture fields over the Wallops Flight Facility and to characterize the multifrequency radiometric signature of tropical convection over the gulf stream and southeastern Atlantic Ocean. A 118-GHz CAMEX spectral data set gathered September–October 1993 was evaluated by the neural network cell-top estimator and the results were plotted. The altitudes produced by the network show ex-

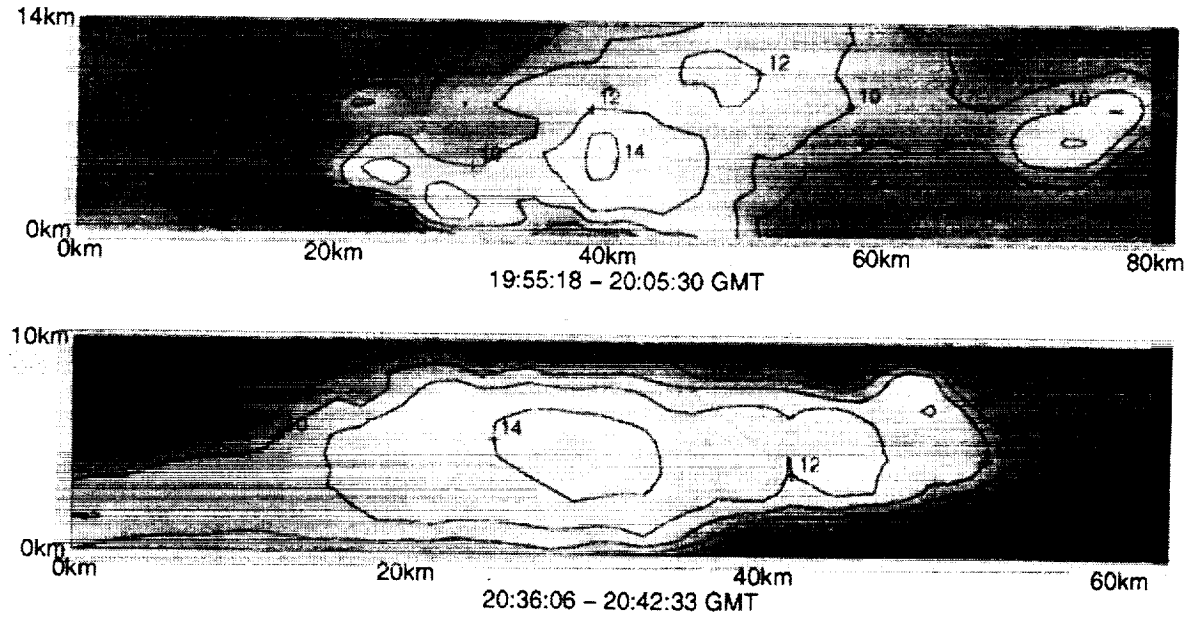


Fig. 4. Images of retrieved precipitating cell-top altitude from CAMEX using a neural network with a single hidden layer.

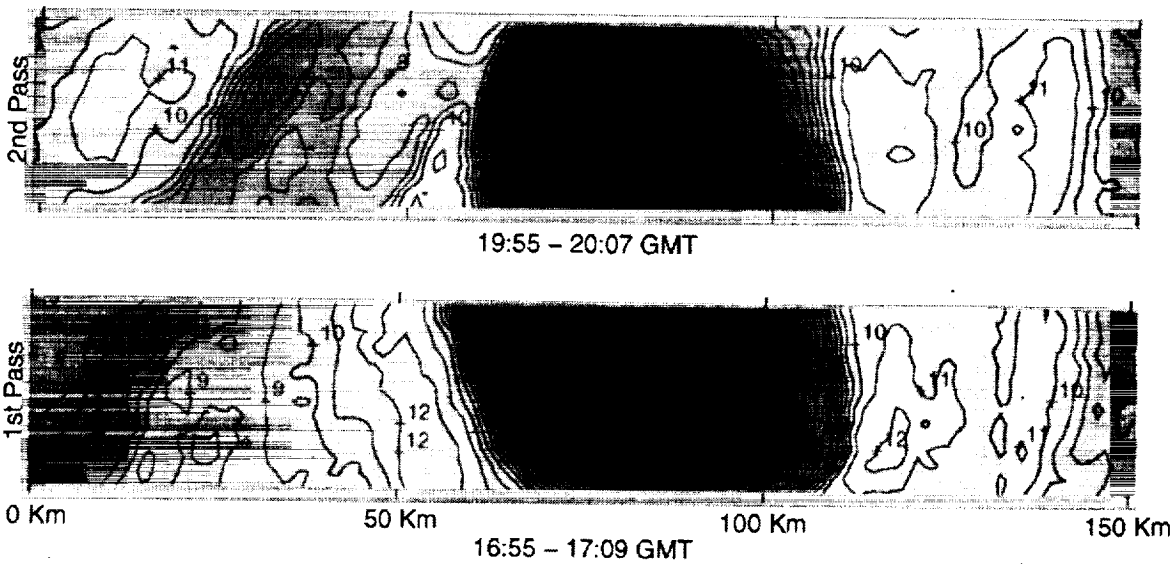


Fig. 5. Images of retrieved cell-top altitudes from Cyclone Oliver using a neural network with a single hidden layer.

pected cell morphology. Fig. 4 shows two samples of imagery of the retrieved cell-top altitudes.

The accuracy of the images was determined by optical estimation of cell-top altitude from the CAMEX video imagery. For near-nadir scan angles, the *a priori* variance in the CAMEX cell-top altitude data is only 1.35 km. The network produced an rms error of 0.84 km, which is an improvement of 38%. Off-nadir scan angles, having an *a-priori* variance of 1.29 km, produced an rms error of 1.19 km. This is an improvement of nearly 8%. The decrease in performance of the neural network estimator for off-nadir scan angles can be explained in two ways. First, the neural network was trained on near-nadir GALE and COHMEX data. Therefore, it is expected that the network will produce more accurate results with similar data. Second, the video imagery of the CAMEX flight showed that the distinct cell-top peaks occurred primarily at near-nadir scan angles, and the off-nadir scan angles showed an increased amount of cirrus cover. In general, the CAMEX pilots tried to fly directly over the most intense and well-defined precipitating cells. Therefore, the off- nadir optical estimates of cell-top altitude were typically higher than indicated by the 118-GHz spectral data.

A second example of the utility of the neural network cell-top estimator is shown in Fig. 5. To make the two images commensurate, the axes were reversed for the second pass so that the N, S, E, and W directions are the same for both images. Cell-top altitude retrievals of Cyclone Oliver (February 7, 1993) over the Pacific Ocean just north of Australia were performed using the neural network estimator developed for data set 1. Although the retrievals could not be

verified with video imagery because the 118-GHz data was gathered at night, the retrieved images reveal the expected general morphology of the cyclone. The absolute accuracy of the retrievals from Cyclone Oliver may be less than that of the CAMEX retrievals because the network was not trained on tropical data. These considerations aside, the retrieved images are still quite useful. The morphology of the eye-wall and the surrounding precipitating cells are clearly visible.

IV. Concusions

Retrieval of cell-top altitude from 118-GHz perturbation spectra has been demonstrated from MTS observations during GALE and COHMEX, using a multilayer, feedforward neural network trained with the backpropagation algorithm. The neural network retrieval method yields an rms error of 1.76 km for a data set consisting of mature, cumulus and dissipating cells. The rms error is reduced to 1.44 km when only cumulus cells are considered. When compared to linear and nonlinear retrieval methods on the same data, the neural network yielded superior results. This may be partly attributed to the fact that the neural network is able to capture not only the nonlinear physical relationship that exists between the 118-GHz brightness temperatures and cell-top altitude, but also the complex statistics of the 118-GHz data.

Improvement in the 118-GHz retrieval is obtained by incorporation of auxiliary 118-GHz observations of atmospheric temperature and cell size into the neural network estimator. With the use of an incremental training algorithm to reduce the complexity of the network when incorporating these additional inputs, the rms error is reduced to 1.36 km.

Finally, it was shown that the neural network estimator could be used to produce cell-top altitude imagery that displays cell morphology in a useful way.

Acknowledgment

The neural network cell-top altitude estimation system was developed on a 486-based PC using MATLAB, version 4.0, and the MATLAB Neural Network Toolbox, version 1.0. The authors would like to thank J. Barrett for his help with the spectral observations and P. Rosenkranz for his helpful discussions.

References

[1] R. F. Adler and R. A. Mack, "Thunderstorm cloud height-rainfall rate relation for use with satellite rainfall estimation techniques," *J. Appl. Meteorol.*, vol. 23, pp. 280–296, 1984.

[2] C. Cabrera-Mercader, "Neural network statistical retrieval of atmospheric water vapor from microwave radiometric observations," S.M. thesis, Mass. Inst. Technol., Cambridge, 1993.

[3] D. T. Davis, Z. Chen, L. Tsang, J. N. Hwang, and A. T. C. Chang, "Retrieval of snow parameters by iterative inversion of a neural network," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 842–852, July 1993.

[4] M. S. Dawson and A. K. Fung, "Neural networks and their applications to parameter retrieval and classification," *IEEE Geosci. Remote Sensing Soc. Newsletter*, pp. 6–14, Sept. 1993.

[5] R. A. Dirks, J. P. Kuettner, and J. A. Moore, "Genesis of Atlantic lows experiment (GALE): An overview," *Bull. Am. Meteorol. Soc.*, vol. 69, pp. 148–160, 1988.

[6] A. Gagin, D. Rosenfeld, and R. E. Lopez, "The relationship between height and precipitation characteristics of summertime convective cells in south Florida," *J. Atmos. Sci.*, vol. 42, pp. 84–94, 1985.

[7] A. Gallant and H. White, "There exists a neural network that does not make avoidable mistakes," in *Proc. IJCNN 88*, vol. I, pp. 657–664.

[8] A. J. Gasiewski, J. W. Barrett, P. G. Bonanni, and D. H. Staelin, "Aircraft-based radiometric imaging of tropospheric temperature and precipitation using the 118.75-GHz oxygen resonance," *J. Appl. Meteorol.*, vol. 29, no. 7, pp. 620–632, July 1990.

[9] A. J. Gasiewski and D. H. Staelin, "Statistical precipitation cell parameter estimation using passive 118-GHz O_2 observations," *J. Geophys. Res.*, vol. 95, no. D15, pp. 18 367–18 378, 1989.

[10] V. L. Griffin, A. R. Guillory, M. Susko, and J. E. Arnold, "Operations summary for the convection and moisture experiment (CAMEX)," NASA Tech. Memo. NASA TM-108445, NASA Marshall Space Flight Center, Mar., 1994.

[11] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with Gaussian hidden units as universal approximations," *Neural Comput.*, vol. 2, no. 2, pp. 210–215, Summer 1990.

[12] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[13] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Acoust., Speech, Signal Processing Mag.*, vol. 4, pp. 4–22, Apr. 1987.

[14] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proc. Int. Joint Conf. Neural Networks*, July 1990, vol. 3, pp. 21–26.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, vol. 1, 1986, pp. 318–362.

[16] W. L. Smith and C. M. R. Platt, "Comparison of satellite deduced cloud heights with indications from radiosonde and ground-based laser measurements," *J. Appl. Meteorol.*, vol. 17, pp. 1796–1802, 1978.

[17] R. W. Spencer and D. A. Santek, "Measuring the global distribution of intense convection over land with passive microwave radiometry," *J. Clim. Appl. Meteorol.*, vol. 24, pp. 860–864, 1985.

[18] M. S. Spina, "Application of multilayer feedforward neural networks to precipitation cell-top altitude estimation," S.M. thesis, Mass. Inst. Technol., Cambridge, 1994.

[19] L. Tsang, Z. Chen, S. Oh, R. J. Marks, II, and A. T. C. Chang, "Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 1015–1024, Sept. 1992.

[20] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I.* New York: Wiley, 1968, pp. 174–187.

[21] T. T. Wilheit, "Some comments on passive microwave measurements of rain," *Bull. Am. Meteorol. Soc.*, vol. 67, pp. 1226–1232, 1986.

[22] T. T. Wilheit, A. T. C. Chang, M. S. V. Rao, E. B. Rodgers, and J. S. Theon, "A satellite technique for quantitatively mapping rainfall rates over the oceans," *J. Appl. Meteorol.*, vol. 16, pp. 551–560, 1977.

[23] S. F. Williams, H. M. Goodman, K. R. Knupp, and J. E. Arnold, "SPACE/COHMEX data inventory document," NASA Rep. TM-4006, 1987.

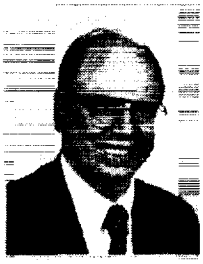
Michelle S. Spina received the B.S. degree in electrical engineering from the Rochester Institute of Technology (RIT), Rochester, NY, in 1991 and the S.M. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1994. She is currently pursuing the Ph.D. degree in the Spoken Languages Systems Group at the MIT Laboratory for Computer Science.

She participated in the cooperative education program at RIT and worked at the AC Rochester Division of General Motors. She also was a Software Engineer in the digital video image transmission systems group at the RF Communications Division of Harris Corporation, Rochester. She also spent a summer at the Cambridge Research Laboratory, Digital Equipment Corporation, working on an audio segmentation system. Her research interests include spoken language systems, speech recognition and understanding, and biomedical issues of speech processing as they relate to automatic speech recognition.

Ms. Spina was a 1995 Intel Foundation Graduate Fellow and is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.

Michael J. Schwartz attended Deep Springs College, Deep Springs, CA, and received the B.A. degree (cum laude) in physics from Carleton College, Northfield, MN, in 1985. He received the Ph.D. degree in physics from Massachusetts Institute of Technology (MIT), Cambridge, in 1997.

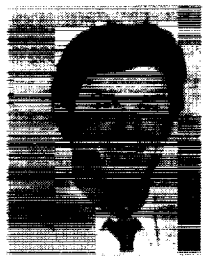
He is currently a Staff Scientist in the MIT Research Laboratory of Electronics and is involved in the development of the NAST-M aircraft-borne, oxygen-line, millimeter-wave radiometer package. His research has concerned aircraft-based measurements of atmospheric microwave opacity due to diatomic oxygen.



David H. Staelin (S'59–M'65–SM'75–F'79) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1960, 1961, and 1965, respectively.

He joined the MIT faculty in 1965 and is Professor of Electrical Engineering and Assistant Director of the Lincoln Laboratory. He teaches electromagnetics and signal processing. He was Principal Investigator for the NEMS and SCAMS microwave experiments on the NASA Nimbus-5 and Nimbus-6 satellites and a Co-investigator for the NASA AIRS/AMSU sounding experiment for EOS, the SMMR microwave experiment on Nimbus 7, and the Planetary Radio Astronomy Experiment on Voyager 1 and 2. Other research has involved radio astronomy, video coding, milliarc-second optical astrometry, and manufacturing process characterization.

Dr. Staelin is a Fellow of the AAAS.



Albin J. Gasiewski (S'81–M'88–SM'95) received the B.S. degrees in electrical engineering and mathematics and the M.S. degree in electrical engineering from Case Western Reserve University, Cleveland, OH, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1989.

Since 1989, he has been a faculty member at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta. His technical interests include passive and active remote sensing, radiative transfer theory, electromagnetics, antennas and microwave circuits, electronic instrumentation, meteorology, and oceanography. His current research concerns remote sensing of the Earth's atmosphere and oceans using passive microwave sensors. During the summer of 1996, he was a Visiting Scientist at the NOAA Environmental Technology Laboratory, Boulder, CO.

Dr. Gasiewski is a member of the IEEE Geoscience and Remote Sensing Society's Administrative Committee, where has been active in supporting GRS chapter development and symposia. He is also a member of Eta Kappa Nu, Tau Beta Pi, URSI (Commission F), Sigma XI, American Meteorological Society, COSPAR, and the American Geophysical Union. He served on the U.S. National Research Council's Committee on Radio Frequencies (CORF) from 1989 to 1995.